Chapter 5

Health Effects Assessment

Learning objectives

- Understand the spectrum of potential adverse environmental health effects in terms of type, severity, timing of onset and duration.
- Be able to develop a case definition and symptom questionnaire.
- Understand the strengths and weaknesses of various data sources for health outcomes including vital records, survey data, hospital records, and registry data.
- Understand typical environmental health effects in key organ systems and be aware of sources of information on health effects of specific exposures.

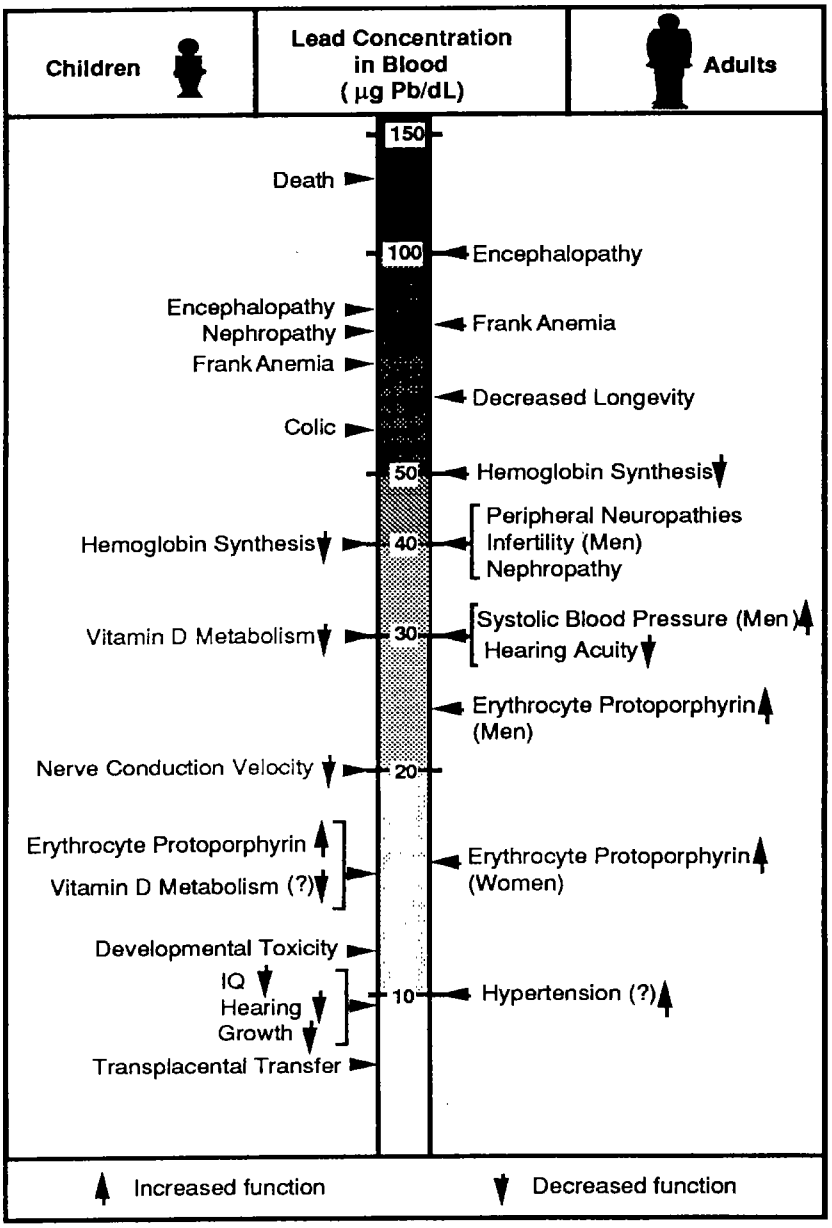
5.1 Concepts of health effects measurement

All organs and systems of the body can be affected adversely by exposures to environmental hazards. Adverse effects range from subtle physiological and biochemical changes that may be asymptomatic, to individual perceptions or symptoms of illness, to clinically diagnosed disease, and finally, to death. In general, toxic agents have specific molecular and cellular components which they target – so that many toxic agents cause specific effects in particular target organs while other agents impact on multiple tissues and organs of the body.

Toxicity is the capacity of a toxic agent to produce injury in an organism. Severity of toxicity depends on the route and magnitude of exposure and on the dose received by the target organ system. The extent and pattern of injury at a given dose is modified by route of absorption and, in the case of chemicals, by the distribution and metabolism in the organism, and the rate of excretion. The severity of toxicity is also affected by the extent to which the organism is susceptible to the hazard. Consequently, even with a similar dose, toxic effects can vary between humans and other organisms, among human sub-populations such as adults compared to children, and among individuals within the same sub-population.

Lead is an example of an environmental hazard that produces a wide range of adverse outcomes at different doses, including both clinical and subclinical toxicity, and produces different effects at different concentrations in children compared to adults (ATSDR, 1997; Fischbein, 1998). (See Figure 5.1.) **Subclinical toxicity** refers to harmful health effects that may have been caused by an environmental exposure to a toxic agent but are not clinically recognizable. Encephalopathy, anaemia, wrist drop and renal failure are among the clinically obvious manifestations observable at the upper end of the range of toxicity. Slowed nerve conduction, impaired biosynthesis of haem, and altered excretion of uric acid are some subclinical effects. It is important to note that these subclinical changes represent adverse reactions to lead and are not merely physiological adjustments to its presence.

Figure 5.1. Effects of lead in adults and children at different blood lead concentrations



Source: ATSDR, 1992a.

Exposure to organophosphate pesticides (OP's) provides another example of the range of clinical manifestations of toxicity. In acute OP poisoning, symptoms and signs of poisoning can vary from respiratory distress to diarrhea, nausea and vomiting; all of which could be confused with an acute infectious disease process (Baker & Williamson, 1990; Schenker et al., 1998). However, the presence of blurred vision, slowing of the heart rate, pin point pupils, and muscle fasciculations point to a chemical etiology. These findings in association with a history of recent pesticide exposure lead to a tentative clinical diagnosis of OP toxicity. Laboratory measures of

cholinesterase activity (red blood cell or plasma cholinesterase depression) may be used to verify the clinical diagnosis and monitor recovery from the poisoning.

The concepts of health effects caused by environmental hazards are introduced in Section 1.3, page 9 and discussed more in Section 2.2, pp. 24. When developing strategies to measure health effects in an epidemiological study, it is relevant to understand the concepts discussed in the earlier chapters - such as local and systematic effects, reversible and irreversible effects, acute and chronic effects, and immediate and delayed (latent) effects. The reader should refer to Chapter 2 to review these concepts.

Spectrum of effects

Adverse health outcomes include a wide spectrum of health effects, ranging from sub-clinical to death. These effects include biochemical or physiological changes that may or may not be associated with clinical disease (e.g., depressed cholinesterase following exposure to organo-phosphate pesticides, chromosomal markers such as DNA adducts, and changes in airway flow rates on lung function tests), but occur in response to environmental exposures; reproductive and developmental effects (e.g. disorders in embryonic and fetal development); well-defined acute diseases (e.g., diarrhea following ingestion of bacterial-contaminated water, respiratory distress and coma following exposure to pesticides); and, finally, chronic diseases (e.g. cancers related to chemical exposures, such as benzene and leukemia; asbestos and malignant mesothelioma; arsenic and respiratory cancer; radon and lung cancer).

A principal reason to study subclinical changes is that such changes may occur earlier than and be predictive of the future development of clinical disease, leading to strategies for surveillance and prevention of the disease. Subclinical changes are also studied because these changes may provide insight about the mechanisms by which the environmental hazard causes the disease. Finally, at any specific dosage level, subclinical changes may occur in a larger proportion of the population than overt clinical disease so that epidemiological studies of subclinical effects may have greater statistical power or ability to detect an association with the hazard.

Disease categories commonly related to environmental agents are shown in Table 2.1, page 19. Section 5.4 provides an overview of health effects for various organ systems.

Relevance of genetics to environmental health endpoints

Traditionally investigators distinguished between "genetic" and "environmental" causes of health effects; however, this distinction is not appropriate because of the close interaction between genetic and environmental factors (Eaton et al., 1998; Khoury et al., 1994). Genetic damage includes chromosomal aberrations as well as alterations in specific genes. Genetic damage to somatic and germ cells may be the underlying mechanism responsible for many types of health effects including reproductive failure and cancer. Many chemicals can induce chromosome aberrations in somatic cells (Ames, 1979; Huika et al., 1990; Rosenkrantz, 1998). These include benzene, pesticides, nitrosomines, vinyl chloride monomer and a number of other industrial chemicals and drugs. Accordingly, much effort has been made to quantify genetic damage in humans as a health outcome in itself.

Another reason that genetic measurements are relevant in environmental epidemiology is that genetic variation in the population may cause variation in susceptibility to specific diseases. For example, some forms of a polymorphic gene may increase an individual's risk of colon cancer if he or she is exposed to food-borne heterocyclic amines (Lang et al., 1994). Historically, family studies have been used to investigate susceptibility, but new molecular technologies have also made it possible to evaluate genetic susceptibility using blood collection from participants. Candidate polymorphic genes can be identified and the rates of each allelic form measured in the study population. This is a rapidly growing field owing to numerous technological advances (Perera, 1991; Schulte & Perera, 1993). An increasing proportion of environmental epidemiology research studies now include genetic measurements in order to examine the interaction between gene and environment.

Acute and chronic effects (timing and latency)

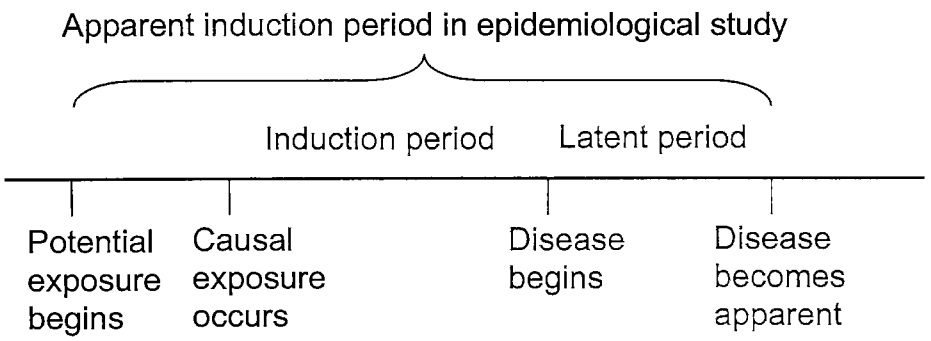
Damage to human health by chemical or biologic hazards is generally related to the dose. In the short term, **acute toxicity** may be defined as hazard dose associated with an adverse health effect occurring in a time frame of minutes, hours or days. In this context, a single dose of the hazard in question may exert toxicity within a few minutes (e.g., acid release cyanide in metal working fluids), or hours (e.g., chemical toxicants requiring metabolic activation) or days (e.g., biologic toxins developed in association with staphylococcal infection).

With some hazards, adverse health effects are delayed. Mechanistically, **delayed toxicity** can be related to a wide variety of factors. Neural tissue damage and behavioural sequella arising from acute carbon monoxide intoxication may be expressed several weeks after the initial insult. However, effects of repeated exposure to asbestos such as development of lung cancer or lung disease (asbestosis) may take many years, even decades, before becoming clinically apparent.

The delay before appearance of disease is generally called the **induction** or **latency (or latent) period**, although these terms have different precise meanings. The **induction period** is the time interval from the causal exposure to initiation of disease, while the **latent period** is the time interval from disease initiation to the clinical manifestation of disease. (See Figure 5.2.) Because the exact causal exposure cannot be determined, the induction period is generally considered the time interval beginning with onset of exposure. Furthermore, because it is difficult to determine exactly the time of disease initiation for most chronic diseases, it may be difficult to distinguish in practical terms between the induction and latent periods. In actuality, epidemiologists tend to use the terms - induction period and latent period - interchangeably to mean the time interval between onset of exposure and clinical manifestation of disease.

For biological agents, the **incubation period** represents the period of time between exposure and the manifestation of disease. During the incubation period, the bacteria, viruses or parasites grow in number and are transported to the critical organ, in which the disease will develop. This period may be hours, days, weeks, or decades in length. For some bacteria, the damage to the critical organ is due to the production of chemical toxins by the bacteria. The incubation period for a biological agent has a different technical definition than the latency period for a chemically induced disease, but in epidemiological studies these terms have the same operational definitions (time between exposure and manifestation of disease) and represent the same concept.

Figure 5.2. Induction and latent periods for a chronic disease



Specific and non-specific effects

Environmental epidemiology studies can be designed to measure a specific outcome related to an identified exposure, or the design can encompass an array of health outcomes within a more general study. The degree of specificity in the health effects assessment will generally depend on existing knowledge about the epidemiological and biological processes, and the quality of the exposure monitoring information for the putative environmental hazard. Studies may focus on one primary, specific effect, but collect information on other secondary or non-specific effects as well. For example, large population-based studies designed to evaluate the effects of chemical air pollutants on respiratory lung disease (e.g., lung cancer) may also be designed to observe the effect of allergens in the study population (reduced FVC, FEV_1).

Host variation

The likelihood of developing an adverse health effect following exposure to an environmental hazard is often affected by characteristics of the persons in the exposed population. The most fundamental consideration is whether the persons are "at risk" of developing a health outcome due to an environmental hazard (see Section 3.1, page 41). Persons may not be at risk of developing a health outcome for gender-specific diseases depending on their gender (e.g. prostate cancer can only affect males), if they already have a health outcome, or if they had a disease and cannot develop the disease again (e.g. immunity following some infections). Even among the at risk population, there is usually variation in the degree of risk depending on characteristics of the persons. General host factors that should be taken into consideration include age, ethnicity, behaviour (such as smoking and amount of physical exercise), hygiene, diet, and co-existing health conditions.

Susceptibility

As mentioned above, some people may have a genetic predisposition for certain effects or increased **susceptibility** to specific environmental hazards. For example, the nevoid basal cell carcinoma syndrome predisposes to radiation-induced skin cancer. In people with this syndrome, the development of basal cell carcinoma is accentuated, occurring only months after the most innocuous X-ray therapy. Similar genetic susceptibility is suspected for some chemicals due to

individual variation in metabolic pathways or enzyme systems (e.g. aromatic hydrocarbons). Another example of susceptibility is the enzyme, glutathione-S-transferase (GST), that functions in metabolism to detoxify absorbed chemicals. This family of isoenzymes catalyses the conjugation of reactive intermediates, such as epoxides formed during early stage metabolism with glutathione scavengers occurring naturally in the body. Those individuals with a low capacity for producing the enzyme have a reduced capacity for conjugating electrophiles, and are more at risk for a mutational event that could be associated with cancer. The increased frequency observed in lung cancer patients indicates that this deficiency is associated with cancer susceptibility in humans (Seidegard et al., 1986). In addition, this deficiency has been shown to be associated with susceptibility to increased chromosome damage by certain agents (Wiencke et al., 1990).

Susceptibility is a particularly important factor to be aware of in environmental epidemiology. As discussed in Chapter 3, relationships between exposure and biological effect or disease can be modified in that portion of a population that is susceptible, and the association between exposure and disease for the susceptible sub-population can be masked or weakened in any study that does not consider variation in individual susceptibilities. Use of susceptibility biomarkers could improve the precision and strength of putative exposure-disease associations by avoiding the dilution effect that occurs in populations with a large proportion of non-susceptible persons (Wilcosky & Griffith, 1990). Consequently, susceptibility biomarkers should be used in stratifying the study populations so that an estimate of disease risk can be made more accurately.

Figure 2.1 on page 26 shows that many potential susceptibility markers may interact in the exposure-disease continuum at several points. For example, an alteration in the cellular membrane permeability could affect the dose through the enhancement of the passage of xenobiotic agents by altering absorbed dose. The induction of carrier proteins could interact with sensitive tissue or cellular targets to influence dosage absorbed by target tissues. The level of repair enzymes operating at the cellular target may modulate the biological effect, thereby altering the severity of the disease. And, alterations in metabolic processes could affect disease progression. For example, N-acetyltransferase (NAT), a non-inducible enzyme that functions in detoxifying aromatic amines such as benzidine and 4-amino biphenyl by acetylation. Two phenotypes are in human populations, one with low acetyltransferase activity (slow acetylators), and one with high enzyme activity (fast acetylators). Those individuals who are slow acetylators are at increased risk of bladder cancer because they can not detoxify aromatic amines as quickly as fast acetylators (Lower, 1982).

Hyperreactivity and hypersensitivity

Hyperreactivity and hypersensitivity are similar terms which are also used to identify highly sensitive individuals in the population, but they have different meanings than "susceptibility" as it is described above. In **hyperreactive** persons, the effects of the agent are qualitatively the same as expected, but quantitatively increased. In hyperreactive persons, "normal" effects occur, but at a lower dose than in the majority of the population. It is possible that the variation in susceptibility is due to the interaction of many genetic and other host factors, but in practice, it is not possible to determine exactly why hyperreactive persons are more sensitive than other members of the population.

Hypersensitivity refers to when persons react with "allergic" effects following exposure to a certain substance (allergen). The allergic reaction involves a person's immune system after it

has become sensitized to the substance. These allergic reactions include skin conditions such as eczema and contact dermatitis, as well as respiratory conditions such as asthma and hypersensitivity pneumonitis.

Case definition

The initial step in an environmental epidemiological investigation is to describe the problem by defining the case. A **case definition** is a set of criteria for deciding whether an individual should be classified as having the condition of interest (CDC, 1992). It includes clinical criteria as well as criteria relating to place and time, and type of individual. Ideally, the clinical criteria should be objectively measurable (such as: fever > 38°C; elevated antibody titres; three or more loose bowel movements per day). Using cholera as an example, a confirmed case may be defined as diarrhoea plus a stool culture that has tested positive for *Vibrio cholerae*.

Within the framework of developing a case definition, several questions must be answered:

- How does the disease present clinically?
- How does one define a case in terms of clinical observations and laboratory tests?
- In what particular population is the disease occurring and is there truly an excess?
- What etiological possibilities immediately present themselves in terms of clinical and population observations?

Whenever possible, a case should be defined on the basis of data derived from clinical and laboratory findings (e.g. pneumonitis as seen on X-ray; altered liver function tests), rather than on the basis of symptoms as reported by participants or their families. However, when clinical evidence is not available or is inappropriate, then it becomes necessary to use other means of establishing a case definition. For example, following the contamination of flour with polybrominated biphenyl (PBB) in Michigan, USA, a "PBB syndrome" was defined through the use of questionnaire data (consisting largely of symptoms such as fatigue, sleep disturbance, joint pain, headache). Questionnaire data may also be combined with clinical data to establish the case definition. For example, data obtained from standardized questionnaires, together with lung function tests, have played a vital role in developing a relatively standardized case definition for chronic obstructive pulmonary disease (COPD).

Case definitions for cancer, respiratory, or cardiovascular disease, are generally well understood by health practitioners. However, for diseases with unknown etiologies, often involving either subtle or severe immunologic or neurologic responses that may be related to environmental exposures (e.g., Parkinson's disease and Alzheimer's Disease), case definitions are not so well defined. Furthermore, outcomes that consist of a blend of subtle and non-specific conditions may be difficult to define. For example, it may be difficult to establish a precise case definition when studying the impact of heavy metal exposure on postnatal development in children. Investigators typically measure several indicators of development, such as gestational age at birth; birth weight; postnatal changes in weight and length; the child's ability to roll over, sit up, and walk at appropriate stages of development; and neurobehavioral function at different ages. Analysis of these variables taken together (e.g., factor analysis or other multivariate analysis techniques) can help to develop a case definition to be used in further studies.

Finally, items related to any etiological hypotheses which might later be subject to epidemiologic analysis should not be included in a case definition. It is obviously impossible to test an hypothesized relationship between illness occurrence and a particular toxic exposure if that exposure is inherent to the case definition.

5.2 Health effects measurement techniques

Data sources

Health effects data may be classified as either primary or secondary depending on the reasons why the data were first recorded. **Primary data** are data that have been collected specifically for a particular study, via, for example, questionnaires or medical examinations. **Secondary data** are data that have been collected for purposes other than for a particular study and generally must be abstracted from existing records (Kelsey et al., 1996). The completeness of registration and accuracy of the information that the records contain can vary widely. Therefore, it is important to be aware of the quality of existing records. The use of existing data for epidemiological studies is discussed further in Section 6.2

Existing data sources for epidemiological studies include hospital records, chronic and infectious disease registries, birth defects registries, hospital admissions, outpatient visits, well child clinics, and environmental monitoring programs. Data gathered from well managed sources are useful because they are collected and disseminated in a reasonably standardized fashion. Data collection for existing data may also be divided into direct (also called primary by some investigators) and indirect techniques. Direct data collection is when an investigator records data on study participants, expressly for the purpose of study (e.g., abstracting information on the study participant from a medical record). An indirect approach may be used if cases of a disease to be studied can be identified through an existing data system. The use of death certificate data is an example of such an indirect, secondary data source.

In many countries laws exists which expressly require reporting of morbidity and mortality data relating to specific chronic or infectious diseases. Such laws compel institutions and health care professionals to report cases of selected diseases or mortality to a central agency or registry.

Data linkage

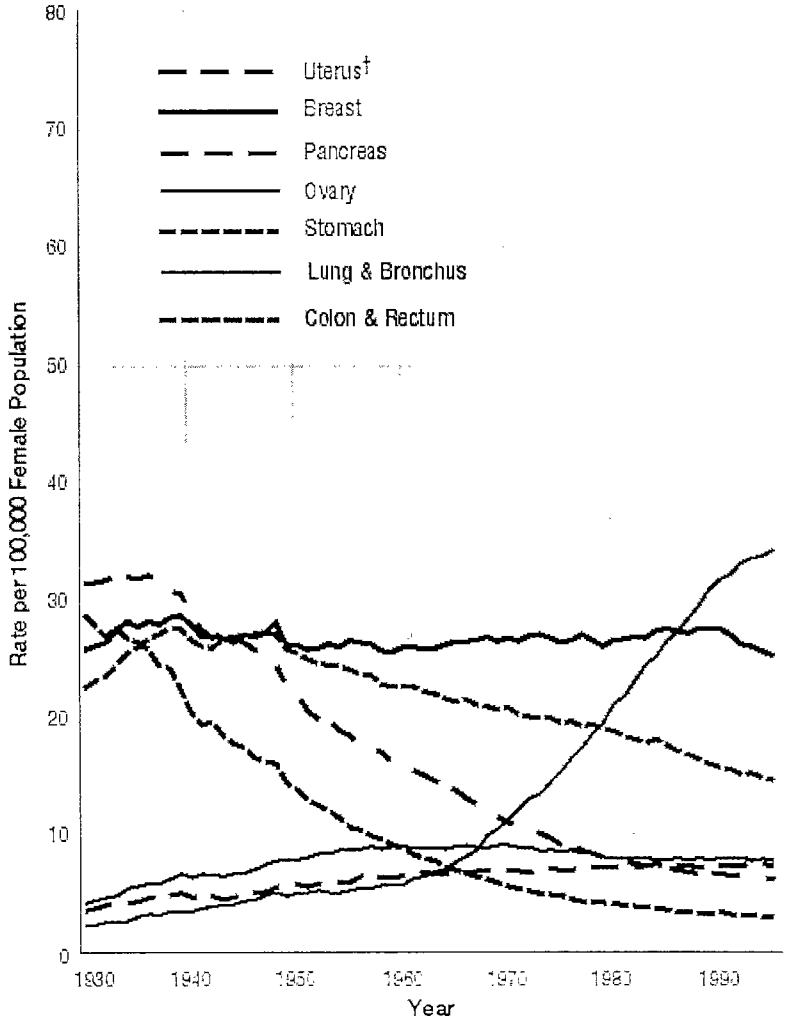
It is beneficial to develop capabilities for linking data sources to provide the best information possible for epidemiological studies. Clearly, the linkage of chronic disease registry data with environmental data bases would facilitate such research. For example, in the United States, efforts are under way by the U.S. Environmental Protection Agency to develop the National Human Exposure Assessment Survey (NHEXAS) and to link data developed in that survey with health and environmental data gathered by other government agencies (e.g., the Agency for Toxic Substances and Disease Registry; and the National Health and Nutrition Examination Survey).

A passive system involving linkage of data bases with information on persons with the disease of interest is frequently used. Although a passive approach is less expensive, it is subject to whatever ascertainment limitations were present in the original data bases. Alternatively, an active approach is used by the Surveillance, Epidemiology and End Results (SEER) registries of the U.S. National Cancer Institute (SEER, 1989). The active approach to data collection

process involves two steps. First, an emphasis is placed on case finding. Then a more detailed data set may be collected for a subset of persons (i.e., those suspected of being exposed or at risk).

In developing data sources, emphasis should be placed on secular (changes over a long period of time) analyses of environmental exposures and disease occurrence. As shown in Figure 5.3, one can see changes that are occurring in female mortality rates for selected cancers over time. Cancer of the lung in females show an upswing in the decade of the 1960's and cancer of the breast in females is also showing a slight increase.

Figure 5.3. Secular changes in female cancer mortality in the USA, 1930-1995



† Uterus cancer death rates are for uterine cervix and uterine corpus combined.
Source: Landis et al, 1999; data from Vital Statistics, 1998.

Mortality and morbidity data

Mortality data

Mortality is influenced by factors that affect duration and survival, such as detection and treatment, as well as incidence. Death certificates are based on the concept of the underlying cause of death, defined as "the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident which produced the fatal injury" (WHO, 1993b). The completeness and reliability of information on mortality varies widely throughout countries of the world. In North America and Europe, for instance, death registration is virtually complete and death certificates generally contain reliable information regarding mortality. While in some other countries and regions of the world, less than 50% of deaths are registered.

The basis for a mortality records system is that all deaths are medically certified by cause, and recorded on a registry that is accessible to appropriate parties. For example, in the United States, researchers can submit names and other identifying information to the National Death Index to determine an individual's vital status and place of death. The appropriate state's Office of Vital Statistics is then contacted for a copy of the death certificate. Other sources of information regarding the occurrence of death include records relating to employment, tax, military service, driver's licence, electoral status and insurance. Professional, union or alumni societies may also yield useful information.

Without accurate clinical data the cause of death may be incorrectly recorded on the registry. Co-morbidity data are rarely available, unless all causes mentioned on the certificate have been coded (immediate and contributory, as well as underlying). Furthermore, in the absence of coding for multiple causes of death, or information collected at autopsy, the cause of death information that is coded from the death certificate may be misleading. Also death certificates generally do not contain information on other risk factors (e.g., tobacco smoking, occupational exposures). Consequently, it is difficult to use mortality data generated from a registry without supplementing those records with other data sources.

It is important to understand that mortality data do not provide information for the investigation of non-fatal health conditions (e.g. asthma, cardiovascular disease, reproductive effects), and may also be unsuitable for studying the etiology of conditions of long duration (e.g. cardiovascular disease), and conditions that have a low case-fatality rate.

Morbidity data

Morbidity data may be gathered from health surveys, medical records (including those kept by hospitals, emergency rooms, clinics and physician's records) and disease registries. Abstracting information of interest from paper records can be very time-consuming. Furthermore, information obtained from medical services only reflects the population of patient who use the specific medical service, and is not likely to be representative of the community population. Data gathered from properly designed epidemiological studies that utilize questionnaire-generated data can be a reliable source of morbidity information.

Hospital records

Hospital records can provide useful information on morbidity and mortality although the investigator must be careful in using such records because of factors associated with admission practices of the hospital (hospital records are not always complete; physicians do not hospitalize

all their patients with a particular disease; some patients die before admission); or because the study participants do not accurately reflect the catchment area for all hospital admissions. For example, if an appreciable proportion of patients admitted to the hospital are from areas outside the community of interest, that hospital's morbidity or mortality statistics will not be reliable as measures of disease prevalence within that community. Clearly, socioeconomic status as well as medical insurance policies may play a role in the hospital admissions process, as well as in the mode of treatment. One should be aware of such issues in deciding whether hospital records are reliable to use for study data.

Employment and school records

Records from work or school may be used to estimate morbidity within selected populations. There are two main limitation in these records. First, the records are applicable only to generally unrepresentative source populations. Second, medical information in the records may not be accurate. In some instances, "sick leave" may be reported for reasons other than sickness. For example, school absenteeism is known to relate to the day of the week, season of the year, and to social events or behavioural factors influencing the student population.

Registries of reportable diseases

When physicians or other health care providers are legally required to report the occurrence of specific disease conditions, such as tuberculosis, congenital malformations, or cancer to a centralized system, the accumulated data-sets are known as registries (e.g. cancer registries). Environmentally related disease incidence data may also be reported to registries (e.g., pesticide poisonings to Poison Control Centre data banks) in some governmental jurisdictions. However, the completeness of reporting is determined by the willingness of the providers to report the cases in a timely and reliable fashion. The validity of such reports can be assessed through the use of community surveys and by checking medical records.

Data generated from birth and birth defects registries are useful for evaluating effects (e.g., infant mortality, reduced birth weight, early fetal loss) related to environmental exposures within communities. However, birth certificates contain only limited information regarding adverse health outcome measures (e.g., the birth record will only record the most severe structural or functional abnormality noted at birth, and not late developing abnormalities), and must be evaluated with care.

Health interview surveys

In the health interview survey, a sample of individuals is questioned about their social setting, their recognition of signs and symptoms of disease, their attitude towards sickness and health, and contact they have had with health services. The population sample may be representative of the total population, or selected subpopulations chosen by geographic distribution, or other characteristics. Such a survey can be a cost-effective means of obtaining data from a large sample of respondents. In comparison, the use of medical and other staff to examine subjects would require far more resources. Data can be obtained from subjects by a variety of methods, including self-completion questionnaires, direct interviews, group interviews, or diaries.

Health examination surveys

In health examination surveys, information is collected by interview, and by physical examination. Problems that may be encountered with this type of survey include the magnitude of resources required to collect the sample and to examine the study population. Clearly, the costs of the physical examinations, combined with the costs of the collection of human tissue and

fluid samples, and the analyses of samples, has a marked influence on the size of the study population. However, the great benefit of the health examination survey is that through the identification of physiological, neurological and psychological parameters, variations from the norm can be compared and related to a standard referent population for comparison.

Questionnaires

Questionnaires are the primary means by which epidemiologists gather data on the characteristics of a study population. Through questionnaires, epidemiologists can gather information on the demographics associated with the study population (e.g., age, sex, race, ethnic group); social factors (e.g., occupation, education, socioeconomic status); personal habits (e.g., smoking, medications, drug use); genetic characteristics (e.g., familial health conditions); and finally, individual and family health histories (e.g., hypertension, diabetes, heart disease, cancer, reproductive effects). As discussed in Chapter 4, questionnaires are also used to collect information on a person's time, activity, and location - which are useful in estimating exposures.

Questionnaires should be designed with brevity in mind, although there is always the temptation to "ask one more question." Often the participation rate is inversely proportional to the size of the questionnaire (i.e., the larger the questionnaire, the poorer the response rate). Questionnaires should be designed to answer very specific questions, if possible prepared so that the respondent must choose from a limited range of predefined answers. For example, questions designed to develop information on miscarriages in a group of female fruit pickers exposed to pesticides are shown below:

1. Have you ever been pregnant? Yes ( ) No ( )
   - 1a. If yes, what is the total number of pregnancies you have had? (specify number)_____
   - 1b. If yes, what is the total number of miscarriages you have had? (specify number)_____
   - 1c. If yes, what is the total number of live births you have had? (specify number) _____

2. Are you pregnant now? Yes ( ) No ( )
   - 2a. If yes, which month of pregnancy are you in now? (number of months of pregnancy)_____

3. Have you had a pregnancy that ended in the past 12 months? Yes ( ) No ( )
   - 3a. If yes, how many months ago did that pregnancy end? (number of months)_____

4. Are you breast feeding? Yes ( ) No ( )

Another example is a group of questions concerning hypertension in a study population:

1.	I have some questions about hypertension (high blood pressure):			
	a.	Have you ever been told by a doctor that you had high blood pressure?	Yes ( )	No ( )
	b.	Another name for high blood pressure is hypertension. Have you ever been told by a doctor that you had hypertension?	Yes ( )	No ( )
	c.	If you have been told by a doctor that you have high blood pressure (hypertension), how long ago were you told? (specify number)	____ months	____ years
	d.	Was it in the past year?	Yes ( )	No ( )

Using the "did a doctor tell you" approach reduces the opportunity for misinformation through improved recall. People are more likely to recall a health event correctly when there has been a physician's diagnosis.

Language and cultural characteristics of population under study should be considered when designing a questionnaire. Questions should be expressed in simple, unambiguous language so that they are readily understood by all members of the study population. The questionnaire should be written in the most common language(s) used by the study population, and the interviewers should be fluent in the language, or have an interpreter available to help with administering the questions. One should consider respondents' possible reaction to questions on the instrument, since the location of certain questions (e.g., example, information on socioeconomic status, education, and income) may lead the respondent to reject further questions.

When designing questionnaires the investigator should ensure that:

- adequate demographic information on study participants is requested;
- questions are properly grouped and are answered so as to reduce the likelihood of misinformation, bias, or non-response;
- skip patterns are clearly stated and easily identified;
- directions for recording answers are clear, space is adequate, and there is a mechanism to separate missing data (non-response) from "zero" data;
- the questions are "understandable" by the study population; keep questions as simple and to the point as possible;
- only "necessary" questions are asked of the respondents;
- the questionnaire is properly "coded" and machine readable, if possible.

In conducting a survey, the investigator must decide whether to use a self-administered questionnaire or to use interviewers to elicit information from the respondents. Although self-administered questionnaires are less costly, the response rate is frequently better with interviewer administered questionnaires. Interviewers are also able to communicate with the participant,

both to allay concerns, and to elicit difficult to derive information. With an interviewer, the respondent can also seek and receive clarification of a question, or help in completing the form. Thus, with an interviewer, there is a better opportunity for reducing bias arising from failure of the respondent to respond entirely, or to provide an incomplete response.

Validity and reliability should be an important consideration when developing a questionnaire (See section 5.3). For example, questionnaires should be designed to minimize potential error by leaving all mathematical calculations to follow-up data analyses. When investigating the age at onset of a given disease, it would be most appropriate to give the year; the age can then be calculated from the birth date.

To test validity within the questionnaire, it is appropriate to utilize an external criterion such as physician's diagnosis of hypertension to validate a question on high blood pressure. As another example, cotinine analyses are commonly used to "validate" questions on smoking, although technically they cannot "validate" smoking history because both cotinine measurement and questionnaire responses have potential for error. That said, the questionnaire might ask:

In the past two weeks, how many packs of cigarettes have you smoked each day?

a.	Do not smoke	_____
b.	< 1 pack per day	_____
c.	1 pack per day?	_____
d.	> 1 pack per day?	_____

The investigator would also analyse cotinine in the participant's urine or saliva to assess consistency with the questionnaire response. (See Table 2.3 on page 26, which shows data from a comparison of smoking history and cotinine concentrations.) However, such biological verification may be expensive and time consuming.

It is also useful to administer the questionnaire to the same person on two or more occasions to evaluate reliability. Unfortunately, in large studies, investigators are not usually able to utilize external measures of validity, or administer questionnaires multiple times. It is sometimes possible to test validity and reliability in a pilot study to gain insight into these issues, prior to undertaking the full study of the study population. The role of a pilot study is discussed further in Section 8.5.

It is possible to construct a questionnaire so that "internal consistency" and reliability can be evaluated. This can be done by asking two or more questions that are designed to elicit a similar answer by the participant. For example, in attempting to gather information on "pesticide safety" and labelling information, one could ask the following question:

Are you familiar with the safety information provided on the pesticide label? Yes ___ No ___

Toward the end of the interview, the following question could also be asked:

What safety information is contained on the pesticide label?

1.	Active ingredients	Yes ___	No ___
2.	Preparation and application information	Yes ___	No ___
3.	Precautionary measures	Yes ___	No ___
4.	Antidotes and first aid instructions	Yes ___	No ___
5.	Storage and disposal instructions	Yes ___	No ___
6.	Toxicity information	Yes ___	No ___
7.	Does not know what is on label	Yes ___	No ___
8.	Other (specify) _____		

Whenever possible, questionnaires should be based upon previously used and tested questionnaires. This has advantages in that design problems may already have been identified and corrected, and is especially useful if data are to be compared with those from previous surveys. For example, the reproductive questions cited above were abstracted from the National Health and Nutrition Examination Survey (NHANES), of the United States, and were used in a study of field workers exposed to pesticides. The reproductive data from the pesticides study were then compared with the NHANES data. An increasing number of investigators are placing their questionnaires or other study instruments on the Internet so that it is becoming an excellent source to obtain existing standard questionnaires.

Physical examinations

It can be challenging to obtain high quality data based on physical examinations even when collaborating with highly qualified clinical examiners. Physicians and other health practitioners are not generally accustomed to perform and interpret physical examinations in a highly uniform manner, which is essential for quality assurance in an epidemiological study. Interpretation of routine clinical findings; qualitative (e.g., skin pallor), and quantitative (e.g., respiratory rate) are variable between clinical specialists. Measures should be taken to minimize variability by clinical observers:

- Establish uniform clinical criteria and clinical review procedures

Prior to beginning a study, establish detailed guidelines for examination and interpretation of the physical findings. The examination procedures should be reviewed with the clinician observers in a random sample of the study population prior to the definitive study. For quality control purposes, one clinician or an appropriate clinical panel should periodically review the clinical data for final interpretation.

- Where possible, employ quantitative clinical measures of pathophysiological effects

Blood pressure, pulse, respiratory rate, and temperature are examples of routine clinical measures which can be most efficiently quantified and calibrated through instrumentation. Less efficient semi-quantitative measures include pupil diameter, extremity reflexes, joint range of motion, gait, liver and spleen size. These semi-quantitative measures depend almost entirely on consensus among participating clinicians. Careful coordination among epidemiologists and the clinical observers is absolutely essential to ensure accuracy and consistency. As discussed in Chapter 8, training even of qualified clinicians and quality assurance review of physical examination procedures are necessary in every study.

Physiological measures

Measurement of physiological effects offers another means of assessing adverse health outcomes. However, because of considerable intra- and inter-variability encountered between individuals, and equipment, every attempt must be made to standardize procedures. For example, in measuring lung function following exposure to airborne pollutants, one sees daily and seasonal variability with specific lung function parameters such as the maximum expiratory volume with maximum effort to full inspiration (i.e., PEFR and FEV_1) (Dockery and Pope, 1994).

Lung function tests (e.g., spirometry) are often used to determine the respiratory health status of subjects participating in air pollution epidemiologic studies. When using spirometry to measure lung functions such as FVC, FEV_1, etc., the comparability of the equipment used, and of its operators will determine whether results from separate studies can be compared. For example, Green and co-workers (1974), in a study of the variability of maximum expiratory flow volume curves, found that flows exceeding 70% of vital capacity varied substantially between individuals, which was attributed to the degree of individual effort expended in completing the procedure. Guidelines have been published to standardize the performance of pulmonary function tests so that findings of epidemiological studies are comparable (ATS, 1995).

The advantages or disadvantages of physiological measurement techniques or instruments have to be judged on the basis of:

- acceptability to the study population;
- the accuracy and reliability of the results obtained using them;
- their ease of use and the availability of technicians who know how to use them.

The sensitivity and specificity of the test must be determined. The determination of what constitutes a true or false positive or negative will depend on the standards employed.

Laboratory testing

Each laboratory test has three properties: sensitivity, specificity and predictive value. The **sensitivity** refers to the ability of the test to identify or detect a biologic or analytic endpoint at some concentration. The **specificity** refers to the ability of the test to identify an effect with

minimal interference from competing reactions or biologic effects. The **predictive value** refers to the ability of the test to determine a relationship to a specific health effect.

Analytic chemical techniques in general are highly sensitive methods designed to quantify concentrations of a given chemical, element or biologic (protein, DNA, lipid). Current analytic technologies are highly specific, although they lack predictive value. For example, knowledge of the concentration of dioxin in tissue alone does not identify whether a subject has or will have an immunodeficiency condition. On the other hand, laboratory assays that detect a specific infectious agent in consistent association with disease outcome have specificity and predictive value with variable sensitivity. For example, use of western blot technology, an immunoassay, to detect AIDS is highly sensitive, relatively specific and is predictive of a long term health consequence of infection. Detection of *Yersinia pestis* by culture of lung fluid at autopsy in association with multiple cases of unexplained lethal pulmonary edema is highly specific with relatively low sensitivity and good predictive value.

Biomarkers

Biological markers can be consider a special type of laboratory test that are being increasingly used in epidemiological studies. The basic principles concerning biomarkers are described in Chapter 2. Biomarkers have been identified for many different types of outcomes. For example, biomarkers of reproductive effects are reduced sperm count, altered sperm morphology, mutagens in body fluids, micronuclei, and somatic cell mutation. Biomarkers of neurologic effect include depressed acetylcholinesterase in serum and plasma, and slowed nerve conduction in peripheral neuropathy. Quantitative measurement of carboxyhemoglobin formation is a marker of carbon monoxide poisoning, but this measurement can give spurious data unless current smoking, and respiratory health status is accounted for (see the discussion of carbon monoxide on pp. 68-69). Sister chromatid exchange (SCE), a cellular measure of chromosome damage and repair, is a sensitive semi-quantitative measure of extent of damage by some genotoxic agents (ethylene oxide, styrene) but is an insensitive marker for a number of other genotoxic agents including benzene. In terms of predictive value, there is no firm evidence at this time that long-term elevated SCE is associated with enhanced cancer risk.

Biomarkers of susceptibility involve genetic or environmentally derived metabolic reactions and variability in response to stimuli that potentially influence the response of the target organ to environmental exposure.

Genetic markers

Markers of genetic damage fall into two general categories: cellular and molecular. The majority of cellular methods used for human studies require culture of human lymphocytes (the relatively easy availability of lymphocytes make them very attractive cell populations for cytogenetic analysis). Since the life span of a lymphocyte may vary from only a few days to as much as 20 years, they may be useful for estimating the effects of long-term chemical exposures, although DNA repair occurs quite rapidly. It has also been suggested that since lymphocytes are in contact with many body tissues, it is possible that they may provide an integrated measurement of chemical exposure (Perera et al., 1987). DNA single strand breaks (SSB) have been postulated to be a good indicator of the carcinogenic or mutagenic potential of a compound.

A DNA migration assay, the Single Cell Gel (SCG) technique, has been developed as a sensitive molecular measure of single strand breaks and alkali labile sites in DNA (Singh et al., 1988).

This assay makes use of the migration of DNA strands after unwinding in an electrophoretic field as a measure of genetic damage resulting from exposure to genetically active agents. The degree of migration is proportional to the amount of damage to the DNA. Advantages of the technique include its sensitivity, the requirement for a very small sample size (a finger prick provides a sufficient blood sample), and the fact that intercellular variation in response can be monitored. The SCG assay has been refined and evaluated with a variety of cell types from a variety of mammalian species, including humans (Singh et al., 1990). Many compounds have been identified that can induce the actual break, such as peroxides, tobacco smoke, etc. More recently, compounds such as aldehydes have been identified that inhibit the repair of DNA SSB through various mechanisms such as decreasing the activity of DNA repair enzymes (Krokan et al., 1990). The importance of an inhibition of SSB would be increasing the odds of a mutation occurring due to lack of repair of a lesion induced by another hazard such as tobacco smoke.

The presence of mutagenic agents in the human body can also be assessed by testing blood or urine samples with sensitive microbial assay systems (Hulka et al., 1990). Mutagenic activity has also been assessed by examining human faeces and breast milk. Indications for chromosome breakage activity can be obtained from short-term lymphocyte cultures from peripheral blood samples. Other indicators for genetic activity include sister-chromatid exchanges in blood cells and morphological abnormalities of spermatozoa.

Although there is great interest in the scientific community in applying the technology of molecular biology to epidemiologic studies, it is equally clear that the interpretation of findings has proved difficult. Consequently, guidelines should be developed by scientists, in conjunction with the medical community and other interested groups, concerning the use and interpretation of biomarkers in epidemiological studies. Because of the difficulties in developing appropriate study designs, the feasibility of biomarkers in general, and molecular biomarkers specifically, must be evaluated through methodological research before they are ready for broad application in epidemiological studies.

5.3 Methodological issues in health effects assessment

Validity and reliability of health outcome measurements

Interpretation of findings from epidemiologic studies depends upon the validity and reliability of the measurements undertaken. The issues of validity and reliability are discussed in Section 3.4, pp. 53. The reader should refer to that section for more information. For a health outcome, validity is characterized by sensitivity (i.e., the probability that a sick person will be classified as sick) and specificity (i.e., the probability that a healthy person will be classified as healthy).

To ensure the validity of symptom questionnaires, the diseases that are being assessed must be defined accurately; the symptoms described should be manifestations of these disease entities. The use of stricter diagnostic criteria can be problematic in that improvements in specificity may reduce sensitivity. If the amounts of disease in different populations are to be compared, it is essential that the levels of sensitivity and specificity do not vary between populations.

The validity of a measurement recorded by an instrument is determined by the accuracy with which it quantifies the effect that it is intended to measure. The reliability of an instrument can

be determined by frequent tests in which everything is the same except the time (test–retest). Some instruments are or can be set up to obtain measurements of duplicate samples and results at the same time (split-half testing). This method is often used in conjunction with questionnaires in which "identical" questions are repeated in several sections.

Intra-individual variability in effects

Humans display distinct variation in personal physiological functions that can be identified and measured (e.g., personal cholinesterase values will vary from hour to hour or day to day, or even from month to month depending on the degree of exposure to either exogenous or endogenous agents). For example, Hayes (1982) has suggested that in normal workers unexposed to organophosphate insecticides, one could expect variations in read blood cell cholinesterase (RBChE) that range from 13% to 25%, and for plasma cholinesterase (PchE) that range from 20% to 23%. Kondo & Ito (1996) studied seasonal variation in the circadian rhythm of lung function of children with asthma. They reported that there were significant seasonal differences in means for the one-second expiratory flow rate (FEV_1), and furthermore, that there were significant seasonal differences in the amplitude of the circadian rhythm of the FEV_1 measurements. So even the range of diurnal variation in a physiological measurement can vary across different seasons. As another example, Nahm et al. (1998) found that there is seasonal variation in IgG4 subclass antibodies to house dust mite in sera from mite-sensitive patients with asthma. The concentrations of allergic antibodies were higher in the summer and lower in the winter (when it is dry), and significantly correlated with the concentration of house dust mite allergen (*Der* F 1) in the patient's bedding. Thus, it is essential to be aware of both random and systematic intra-individual variation when recording biological measurements.

Inter-individual variation

Variation in biological response measured in one individual and compared with another individual is known as inter-individual variation. Inter-individual variation must be taken into account when recording biological measurements. For example, from a review of ten studies comparing OP exposed groups to non-exposed controls (Duncan et al., 1986), it may be seen that the coefficient of variation (CV) of inter-person PChE values ranges from 14.9% to 30.7% among unexposed controls for four different units of measurement. There was no particular trend among measurement methods and the mean CV was 22.7%, with a standard deviation of 5.7%. From these data, it is clear that, while inter-person variation is greater than intra person variation, there exists a wide range of variation in normal subjects.

5.4 Environmental health effects by organ system

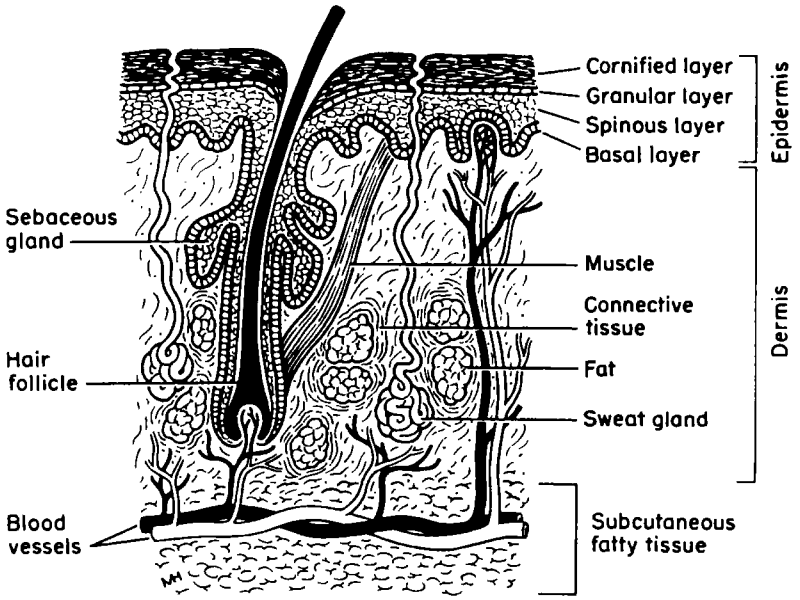
Many sources of information are available that describe health effects caused by environmental hazards. The focus of this book is on epidemiological methods, so only a summary discussion of environmental health effects will be presented. The reader is referred to other sources for more detailed information health effects by organ system (e.g., Rom, 1998; Rosenstock & Cullen, 1994; Yassi et al., 1998). Additional information sources are listed in Chapter 12.

Skin

The skin is a primary target of environmental exposure, as well as a potential route of entry into the body (Adams, 1990). The major components of the skin are the epidermis and the dermis. (See Figure 5.4.) The epidermis includes stratum corneum – the outer layer of densely packed, keratinized epidermal cells that is the dead surface layer of skin, and is the primary deterrent to the absorption of toxic chemicals through the skin. The stratum corneum is replaced about every two weeks in the adult human; a process that involves the dehydration and polymerization of the intracellular material, resulting in the thin, keratin-filled, dried cell, and biologically inactive layer of tissue. The dermis is composed of sweat and sebaceous glands, blood vessels, connective tissue, capillaries, and hair follicles. Blood vessels are distributed throughout the sub-papillary portions of the skin. Toxicants diffuse more readily through the dermis since it contains a porous, non-selective, watery diffusion medium, and toxicants transfuse this tissue by simple diffusion into the systemic circulation, which is dependent on blood flow, interstitial fluid movement, and lymphatic action. Fortunately, the lipid nature of its structure is uniquely developed to form a protective barrier against exposure to and absorption of xenobiotic agents.

For percutaneous absorption to occur, a chemical must pass through the stratum corneum, moving on through the epidermis, through the dermis, and finally, into the systemic circulation. In the human body, the stratum corneum varies in structure and chemistry from one part of the body to another, and this difference is reflected in skin permeability. Thickness of the stratum corneum in the forearm and abdomen is about 8-15 μm, while in the palms and soles it is about 400-600 μm. Toxicants easily cross the stratum corneum of the scrotum since the tissue is very thin, with high diffusivity.

Figure 5.4. Cross-section of the human skin



Source: Wester and Maibach, 1992.

Skin disease

Dermal exposure may lead to acute and chronic effects. Acute symptoms include pain and itching, while signs include, for example, skin rash, erythema, and vesicles. Signs of dermatitis are erythema, vesicles, oozing, and crusting. As the dermatitis becomes chronic, scaling and thickening of the skin occurs.

Contact dermatitis is a commonly reported health problem. If lesions are confined to exposed areas of the body (e.g., hands, face), there is strong likelihood that contact with an environmental agent is responsible. If the lesions are distributed evenly on the exposed surfaces, the possibility of exposure to an airborne agent should be considered. If, on the other hand, the lesions are patchy or linear, one should consider chemical exposure, or perhaps exposure to a plant as the causative agent. In determining the etiologic agent, time of eruption, distribution of lesions, and association of lesions to specific types of exposures should be considered. Allergic mechanisms also cause contact dermatitis. Metals, such as nickel, mercury, and chrome, dyes, resins, and adhesives are common environmental antigens.

Hypomelanosis (depigmentation) results from the loss of melanocytes (unicellular glands that secrete melanin particles into epidermal cells). The principal basis for colouration in the skin, hair, and eyes is melanin. Melanin functions as the screen that shields the dermis from the debilitating effects of solar radiation. Inorganic trivalent arsenic can produce hypomelanosis. The condition is a marker for several clinical conditions (e.g., Addison's disease, chronic protein deficiency, ulcerative colitis, phenylketonuria, and arsenical intoxication). Developing on the mucous membranes of the cheeks, gums, and tongue, primarily in smokers, leukoplakia is a precancerous marker characterized by white, thickened patches of tissue. Actinic keratoses, wart like or horny growths on epidermal tissue, are precancerous markers of squamous cell carcinoma, resulting from excessive exposure to solar radiation.

Respiratory system

The lungs are designed to promote the exchange of oxygen and carbon dioxide. The adult respiratory tract exchanges approximately 5 - 10 m^3 of air daily (Craighead, 1995). The volume and surface area of the respiratory system increases in size from the nose to the alveoli (the alveolar region is the lung area where toxic agents are most readily absorbed). The respiratory surface has been estimated at 50 - 70m^2. Cilia are found throughout the nasopharynx and the tracheobronchial tree. The epithelia tissue carries mucus from the nose to the pharynx, and from the bronchioles through the bronchi and trachea to the pharynx.

Toxic agents may enter the respiratory system as either particles (e.g. dust, pollen), gases (e.g., carbon monoxide, ozone, nitrogen dioxide), vapours of volatile or volatilizable liquids such as benzene and carbon tetrachloride, or aerosols (liquid droplets or fine particles such as those in fumes or smoke).

The pulmonary system is required to process a large and complex mixture of foreign materials that may cause disease in the exposed person. Importantly, when toxic gases and particulates move through the airways, macrophages and neutrophils are generated that create cytokines, proteases, and oxygen species which tend to alter the permeability of the mucosa and play a role in particle and gas uptake by the tissue (Craighead, 1995).

Liquid aerosols can pass through the alveolar cell membranes by diffusion, if the aerosol is lipid soluble. When a toxicant reaches the lungs, it is usually absorbed quite rapidly due to the thinness of the alveolar membrane and the abundant blood supply, and may then be passed through the lower part of the respiratory tract to the target organ site.

The diameter and electrostatic charge of particles greatly influences their movement into and through the respiratory tract. Most large particles (> 30 μm) are not inhaled under normal (non-exercise, light exertion) conditions simply because of their size and weight. Particles that make it into the nasal passages (7 μm - < 30 μm) are removed by the internal nares and nasal passages. The epithelia is quite efficient and will remove about 80% of inhaled particles that manage to pass through the nares. Consequently, only small particles (1 - 5 μm) will reach the bronchiole and alveolar ducts. Even smaller particles (< 1μm) will reach the alveoli of the lungs. Once particles reach the alveoli, they tend to remain there until clearance which takes from several hours to several months. Initially, phagocytes ingest the undissolved particles and through ciliary movement take them in a matter of hours to the pharynx where they can be swallowed. Longer lasting particles may be cleared over a period of weeks, months or years.

Respiratory tract health effects

Inhalation of toxic agent can cause a wide range of effects on the respiratory tract, ranging from non-specific symptoms such as shortness of breath and tightness in the chest, to more specific diseases such as bronchial asthma, emphysema, asbestosis, and adenocarcinoma, among others. Table 5.1 shows examples of the effects of environmental exposures on the respiratory tract.

The physical and chemical properties of pollutants described above are relevant because they determine the site and type of reaction of the respiratory tract to the pollutant. For example, large solid particles are filtered out in the nasal passages, while small particles may reach the distal airways and alveolar spaces. For gases and vapours, water soluble agents tend to be absorbed in the upper airways, while more water insoluble agents penetrate to the alveolar spaces. Therefore, for example, formaldehyde (highly soluble) tends to cause upper respiratory tract irritation, while nitrogen dioxide (less soluble) causes lower respiratory tract effects.

Some respiratory tract reactions occur acutely, while others develop slowly into chronic conditions. Examples of acute reactions include mucosal irritation, increased mucous secretion, reversible air way obstruction, pulmonary edema, haemorrhage, vasculitis, and infection. Examples of chronic reactions include chronic airway obstruction, emphysema, fibrosis, and cancer.

Asthma, which is a disease characterized by hyperreactivity and reversible obstruction of the bronchial airways, is a common condition that appears to be increasing in prevalence throughout the world. In many regions, approximately 8-10% of children have asthma. Asthma is often associated with allergy to one or more environmental agents - particularly mites, cockroaches, animal dander, grasses, and pollen (e.g., Rosenstreich, 1997). Asthma is also exacerbated by exposure to irritants, such as tobacco smoke and air pollutants. The reasons for the apparent worldwide increase in asthma prevalence is a current area of research among epidemiologists.

The investigation of lung diseases relies on the use of standardized questionnaires, the physical examination, and physiological measurement of lung function. Lung function testing, especially spirometry, is a reliable and portable technique to evaluate lung breathing capacity and airflow

movement. Many lung conditions show characteristic patterns on these tests - for example, reduced airflow due to airway obstruction in asthma, bronchitis, and emphysema; or reduced lung volume (restrictive pattern) in lung fibrosis. Chest radiographs and more advanced imaging techniques may be used for non-invasive evaluation of the lung. Finally, laboratory analysis of the blood, such as for concentrations of oxygen and carbon dioxide, may be useful in evaluating the lung's capability for vital gas exchange.

Table 5.1 Respiratory tract outcomes due to environmental exposures

Example of exposures	Type of effect
Formaldehyde, sulfur oxides and other highly soluble irritating gases and vapours	Acute respiratory tract irritation
Acrolein, aldehydes, chlorine, nitrogen oxides, and other insoluble gases	Acute pulmonary tissue injury (bronchiolitis obliterans, pulmonary edema)
Allergic reaction to mites, cockroaches, pollen, multiple chemicals	Asthma
Chronic inhalation of irritating vapours, tobacco smoke	Obstructive lung disease (emphysema, chronic bronchitis)
Chronic inhalation of mineral dusts	Pneumoconiosis (e.g. asbestosis, silicosis) and related types of lung tissue fibrosis
Tuberculosis, Legionella bacterium, and other biological agents	Pneumonia and other respiratory tract infections
Allergic reaction within lung to thermophilic bacteria, fungi, animal proteins	Hypersensitivity pneumonitis
Beryllium	Granulomata in the lung tissue
Respirable particulates in urban areas	Hospitalizations and mortality due to respiratory and cardiovascular causes
Asbestos, radon, tobacco smoke	Lung cancer

Central and peripheral nervous systems

Environmental exposures to neurotoxins through occupational use, accidental chemical spills during distribution or application, improper storage and disposal practices, are a significant risk to the public's health (Aldrich & Griffith, 1993). Neurotoxic agents can be classified according to their activity at the anatomic site. For example, specific toxins may cause damage to the:

- myelin sheath, affecting the oligodendrocytes or Schwann cells.
- selected tissue of the central nervous system (CNS).

- neurons and astrocytes (of the gray matter) resulting from anoxia.
- axons of peripheral neurons.
- synaptic junctions of the neuromuscular system.
- perikarya of the peripheral neurons.

The central nervous system (CNS) is relatively protected from many toxicants by the blood-brain barrier (BBB). The BBB exists in the parenchyma and the choroid plexus of the brain, except for an area of the brain that is particularly sensitive to changes in extracellular fluid components (i.e., the hypothalamic area). The BBB is highly permeable to water, carbon dioxide, and oxygen; slightly permeable to the electrolytes, such as sodium, chloride, and potassium; and almost totally impermeable to metals such as arsenic, sulfur and gold. Although the blood brain barrier does not have absolute control over the passage of toxic materials into the CNS, it does represent a site that is less permeable than most of the other areas of the body.

Neurons in some tissue (e.g., dorsal root ganglia and autonomic ganglia) of the peripheral nervous system (PNS), are susceptible to toxicants that slip through spaces between the epithelial cells. Neurologic damage results from the selective penetration of some toxic chemicals through the protective barriers of the CNS and the PNS to specific anatomic sites. The most serious (and irreversible) damage occurs at the death of the neuron, since the differentiated cells cannot divide and be replaced. When neurons die, other cells with the same function (i.e., fortunately, the nervous system is replete with redundancy of function) may be available to replace the dead cells, or other neurons may acquire the necessary function. If this repair mechanism fails, loss of function will occur according to the severity of the toxic insult. Generally, after neurotoxic insult, some recovery of function occurs. When cell death is not involved, recovery occurs quickly (after the toxic agent is metabolized), or after biological regeneration has taken place. When neuronal death occurs, recovery may take longer, or it may not occur at all.

Neurological system health effects

Examples of acute neurological conditions that can result from high-dose environmental exposures include: encephalopathy as a result of lead exposure, blindness following ingestion of methanol, and seizures and respiratory paralysis due to exposure to organophosphate pesticides. Outbreaks of neurotoxic diseases due to environmental contamination have been well documented. For example, consumption of fish contaminated by methyl mercury in Minamata, Japan, and of fungicide-treated grain in Iraq resulted in blindness and ataxia. Consumption of cooking oil contaminated with tri-o-cresylphosphate (TOCP) in Morocco resulted in spinal cord degeneration and peripheral neuropathy. (Landrigan et al., 1994) Symptoms of pesticide poisoning include eye and skin irritation, miosis, blurring vision, headache, anorexia, nausea, vomiting, increased sweating, increased salivation, diarrhea, abdominal pain, slight bradycardia, ataxia, muscle weakness and twitching, and generalized weakness of respiratory muscles.

CNS involvement is noted by giddiness; anxiety; restlessness; drowsiness; difficulty concentrating; poor recall; confusion; slurred speech; convulsions; coma with absence of reflexes; bursts of slow waves of elevated voltage in EEG, especially on over ventilation; Cheyne-Stokes respirations; depression of respiratory and circulatory centres with dyspnea, cyanosis and fall in blood pressure. In the latter stage of more severe poisonings, symptoms may progress from difficult and laboured breathing to loss of muscle control, convulsions, and possibly death (Griffith & Duncan, 1985).

Clinically apparent symptoms and signs of neurological impairment (e.g. nystagmus, tremor, paraesthesia, focal signs) can often be detected by examination, or by reviewing the medical history of the patient, or by requesting the patient to complete a medical questionnaire. For example, workers exposed to diaminoproprionitrile (DAPN) (a catalyst used in foam rubber production) reported that they were suffering from symptoms of urinary tract infection. An investigation was able to identify the offending process and likely chemical agent. However, a clinical investigation documented that the symptoms were caused by a chemically-induced neurogenic bladder.

Detection of the subtle clinical and sub-clinical neurological effects of a variety of chemicals such as solvents, pesticides and heavy metals may require the skills of a neurologist or specialists in neurobehavioural testing. Testing for peripheral neuropathy, for instance, may necessitate use of devices to assess nerve conduction velocity. Sensory function can be assessed by determining the vibration and temperature threshold. Other tools include electromyography and evoked potentials. Many advances have been made in techniques to test for neuropsychological dysfunction. For example, standard test batteries have been developed and validated using standard populations. Recently many of these tests have been adapted to be administered by computer, for example, testing of reflexes and visual-spatial coordination. The advantages of the computer-based tests is that they require lower training of study staff and they have less variability (and thus better reliability) than staff-administered tests.

Gastro-intestinal system

The primary function of the digestive tract is to convert nutrient sources into useable forms of protein, fat and carbohydrates, vitamins and minerals essential to sustain and maintain the organism. It also functions in the maintenance of electrolyte balance and in the excretion of waste products. Aside from these physiologic functions, the digestive tract is a major source of absorption of toxicants arising from multiple environmental sources. For example, in daily diet, persons may be exposed to trace metal toxicants such as arsenic in water or in shell fish, grain contaminated with fungal toxins, and food products contaminated with pesticides or biologic toxins.

Health effects to be measured can be acute, subacute or chronic. For an acute effect such as diarrhea, identification of an infectious organism or biologic toxin becomes the target for epidemiologic surveillance in an affected population (e.g., salmonella food-borne infection). Chronic health effects such as oesophageal cancer in association with exposure to aflatoxin or cutting oil-associated colon cancer are measurable endpoints. Biomarkers such as measures of p-53 mutations, p-53 protein and chromosome markers can be helpful adjuncts. Tests for occult blood in the stool may be useful for a large scale surveillance program.

Liver

Everything the intestine absorbs passes through the liver for metabolic processing. As such, ingested foodstuffs, drugs and xenobiotics are activated, transformed and detoxified in the liver parenchyma. Similarly, chemicals, either inhaled or absorbed through the skin, find their way to liver cells through a highly vascular network permeating the liver. In the detoxification process, the majority of toxicants are converted to less toxic water soluble waste products for excretion in the kidney or in bile. Some toxicants including alcohol, aflatoxin, industrial solvents, and

N-nitrosamines, however, are converted to their more lethal, disease-producing form. Further, the ability of liver cells to handle toxicants is genetically driven by a series of liver microsomal enzymes including the cytochrome P-450 group accounting for inter-individual variation in susceptibility to toxicant effects. Enzyme activity and pattern of detoxification can be altered by xenobiotics such as polychlorinated biphenyls (PCBs). Finally, the liver is a target organ, source of toxicant effects, and focus of detoxification for the majority of organic chemicals requiring metabolic processing. Infectious agents including hepatitis B and parasitic Schistosomes may play a role in the susceptibility of the liver to toxicant chemical related long-term chronic health effects such as hepatocellular carcinoma, angiosarcoma and cirrhosis.

Clinically observable liver disease often presents as jaundice with attendant signs of liver failure including clouding of consciousness and other neurologic dysfunction. At this stage of liver disease, perhaps 50% or more of normal liver function has been lost. Acute toxicant hepatitis arising from mushroom poisoning or ingestion of certain drugs in a susceptible host and/or infectious hepatitis can produce these clinical effects as an early event, but most liver disease is silent clinically for many years. Detection relies almost entirely on laboratory investigation and analysis. Routine tests including measurement of serum enzyme activities (e.g., aspartate amino transferase, alanine amino transferase and gamma glutamyl transferase), albumin, prothrombin time, and bilirubin levels can be used to detect liver injury prior to clinical signs of disease. In a more sophisticated testing to specify environmental toxicant effects, measurement of P-450 enzyme activity (PCBs and other polychlorinated xenobiotics), detection of P-450 mutations (aflatoxin), analysis of tissue residual toxicant levels (polychlorinated chemicals including PCBs, dioxins) are among the measures that address particular environmental chemical hazards.

Renal system

Some industrial chemicals, pharmaceuticals, herbal preparations, biologic toxins and heavy metals are renal toxicants, either by direct renal toxicity or indirectly through chemical destruction of circulating red blood cells, or through an autoimmune process. The kidneys' primary function is to filter blood and remove waste products. A secondary function is the production of hormones such as erythroprotein. Anatomically, the kidney is partitioned into glomeruli and tubules. The glomeruli are made up of a highly coiled and intertwined capillary beds that serve the function of filtration. Part of the filtrate, intact protein and salts, are returned to the blood. Part of the filtrate, waste products, predominately from protein degradation, are excreted. The renal tubules are composed of epithelium that serve the functions in tubular secretion and tubular reabsorption. Any one or more of these anatomic structures are targets of specific renal toxicants. For example, inorganic mercury salts, lead and cadmium specifically, exert toxicant effects on the proximal tubules of the kidney. The end result is a protein losing nephropathy through failed reabsorption of protein.

Significant exposure to hydrocarbon solvents can lead to generation of antiglomerular basement membrane antibody by an autoimmune process. The end result is an accumulation of waste products in the blood through reduction of glomerular filtration. Tubular necrosis with attendant loss of function is one possible consequence of chlorinated solvent exposure. Similarly, ingestion of ethylene glycol produces renal tubular toxicity with formation of crystalline deposits of urate in the tubular lumen. Because the functional reserve of the kidney is great and because the kidney can compensate for damage by hypertrophy, clinical presentation frequently occurs

late in the course of disease. For these reasons, testing for epidemiological purposes is directed to laboratory detection of abnormalities. Routine samples of urine can be examined for the presence of protein, glucose, abnormal cellular constituents and specific gravity. Measurement of urine specific gravity provides a simple estimate of the ability of renal tubules of the kidney to concentrate or dilute waste products. The presence of protein in urine is abnormal. Specific quantitative detection of the protein ß-2 macroglobulin can be used to determine the extent of renal impairment due to cadmium exposure. Urinary glucose, while routinely used as a screening test for diabetes, is also used to measure subtle changes in renal tubular function.

Microscopic detection of red blood cells, excess white blood cells, and casts made of protein or cellular debris provide laboratory indicators of acute and chronic renal disease. Chronic renal disease can be a factor in the overall chemical toxicity by alteration of the kinetics of toxicant excretion. More detailed estimation of kidney function is divided into examination of glomerular filtration rate by clearance of creatinine and tubular function by measurement of para-amino hippuric acid excretion (PAH). In the epidemiologic setting, creatinine clearance and creatinine levels are used together with analysis of urinary metabolites of chemical toxicants to provide information on the excretion rate of the toxicant. Renal tubular secretion can be measured by PAH, but is not easily performed outside the clinic setting.

Haematopoietic system

Blood, peripheral blood cells and progenitor cells of the bone marrow are components of the haematopoietic system. The primary functions of this system are blood oxygenation (red blood cells), coagulation (platelet, blood clotting factors such as prothrombin, thromboplastin), control and prevention of infection (neutrophils, monocytes, macrophages, lymphocytes), immunity and tumour prevention (lymphocytes, macrophages, monocytes) and allergic response (lymphocytes, mononuclear cells, mast cells, basophils). Each of the formed (cellular) and non-formed elements (e.g., coagulant proteins) are rapid responders to environmental threats to homeostasis. This system is the first line of response of the host organism to biologic and chemical agents.

Ionizing radiation is a well known source of haematopoietic system damage. In the acute phase, radiation damage produces a profound lymphopenia and affects general haematopoiesis. Affected individuals are highly susceptible to infection, perhaps the most common cause of death in significant acute radiation injury. In the long term persons are at risk of developing cancer, particularly, leukemia, and thyroid cancer.

Lead and other heavy metals can induce anaemia. Benzene is a cause of aplastic anaemia with both red blood cell and white blood cell precursors being affected. Oxidative agents such as naphthalene can produce methemoglobinemia. Persons with glucose-6-phosphatase deficiency (G-6-PD) are particularly sensitive to oxidant injury. Long term benzene exposure has been associated with elevated leukemia risk, and 2,4-D (a herbicide) has been associated with risk of non-Hodgkin's lymphoma.

When the oxygen carrying capacity of blood is compromised, as is the case with the production of carboxyhemoglobin by carbon monoxide, and methaemoglobin by nitrites, the result is cellular damage, neuronal death, and finally, if the intoxication is sufficiently severe, coma and death of the individual.

In summary, the haematopoietic system is sensitive to environmental toxicant injury at many levels. It is also an accessible media for environmental toxicant investigation. Measurement of routine laboratory parameters (red cell and white cell counts with morphologic analysis of a peripheral smear) provide the investigator a simple detection system with biologic sensitivity.

Immune system

The immune system is the primary host defence against infection, tumour formation and removal of foreign materials. It is a mobile defence based on cellular response and humoral antibody-forming components capable of constant search and protection of each organ system. In terms of health effects, environmental toxicants may up-regulate the system (immunostimulation), for example, grain dust, or down-regulate the system (immunosuppression), for example dioxins. They may also produce immune dysregulation in the form of allergy, for example, toluene diisocyante (TDI), autoimmunity, mercury salts, and granulomatous tissue response, for example, silicosis and berylliosis. More subtle forms of immune dysregulation involve shifts in the proportion of cellular components (T-lymphocytes and B-lymphocytes) and specific lymphocyte subsets (helper/killer T-lymphocytes) thereby impairing the ability of the organism to respond to infection or to tumorigenesis (e.g., AIDS). The likelihood or risk of adverse health effects due to immunotoxicants is moderated by the personal health (eg., smoking) and work habits (use of personal protective gear) and genetic background of the individual. For example, atopic persons (i.e., persons with a family history of allergic disorders) are at enhanced risk for asthma given an aerosol exposure to allergic immunotoxicants (e.g., dusts, moulds).

Dermal contact with allergic immunotoxicants (eg., metals such as nickel or some sun screens containing paraben) will lead to enhanced risk for allergic contact dermatitis. Similarly, persons with genetically different immunologic tissue types (HLA) vary by ethnic group and may vary in susceptibility to autoimmune diseases (e.g., ankylosing spondylitis, a form of arthritis).

Laboratory support for investigations of immunotoxicity can be quite complex involving quantisation of T and B cell subsets, tissue typing, and *in vitro* cellular response to immunogens. More routine examinations can include white blood cell count with differential, quantitative measurement of immunoglobulin levels and clinical challenge tests of potential immunogens.

Reproductive system

Scientists believed for years that the developing embryo was well secured from environmental insult since it is protected by the walls of the uterus. The adverse effects of intrauterine radiation on the developing fetus were well-documented by the early 1930's, and by 1941 the rubella virus was observed to cause congenital malformations among the offspring of mothers who had been exposed during pregnancy (Gregg, 1941). In 1961, the thalidomide tragedy demonstrated that the developing fetus was not free from potential insult from exogenous agents (McBride, 1961; Lenz, 1962). With the discovery of a cluster of malformations of the central nervous system in newborns in Minamata Bay, Japan in the 1950's the link between environmental contamination and adverse reproductive outcomes was established (WHO, 1990). Concern over the potential reproductive effects associated with environmental exposures has increased in the last decade with exposure incidents in Seveso, Italy (Lancet, 1981; Bertazzi et al., 1992) and Love Canal

(Vianna & Polan, 1984). Although adverse reproductive effects were believed to be related to these incidents, no clear connection between exposure and health outcomes could be established.

The enhanced susceptibility of the developing fetus and young infant to environmental exposures is another concern. This is due primarily to the rapid rate of cellular differentiation and growth, and the development of complex structures such as the central nervous system. Environmental agents may exert their effects at several points during the reproductive cycle, prior to or during conception and fertilization, prior to implantation, during embryogenesis (first trimester), or during the period of later fetal growth and development (second and third trimesters). The effects of environmental insults on the newborn infant and child can result in a lifetime of disease, disability, and decreased productivity, placing an increasing burden on already strained health resources. In addition, to the clinically evident morbidity, an unknown number of adverse effects results from chronic low level exposures to certain agents. Much of this childhood morbidity may be preventable through early recognition and remediation of hazardous exposures.

The possible adverse reproductive effects that may potentially result from environmental exposures are varied, and range in severity from mild, transient effects (such as temporary menorrhea) to highly lethal conditions (such as major chromosomal disorders). Table 5.2 presents some of the putative adverse outcomes that may be observed. The list is not a complete compilation of all possible effects, rather it presents some of the more clinically significant outcomes which could occur as a result of exposure to environmental factors.

Table 5.2. Potential adverse reproductive outcomes

1. Delayed conception/infertility (sperm abnormalities, ovulatory or menstrual disorders).
2. Sexual dysfunction (decrease libido, impotence).
3. Spontaneous abortion (pregnancy loss prior to 28 weeks gestation).
4. Fetal mortality (fetal loss after 28 weeks gestation).
5. Neonatal mortality (infant deaths up to 28 days of age).
6. Low birth weight (preterm delivery, fetal growth retardation).
7. Single gene mutations (dominant lethal mutations, Mendelian disorders, other phenotypically manifest conditions).
8. Chromosomal abnormalities/aberrations (aneuploidy, nondisjunctive disorders, breaks, deletions).
9. Congenital malformations (single defects, multiple malformation syndromes or associations).
10. Developmental disabilities.
11. Childhood cancers.

Source: Aldrich & Griffith, 1993.

Environmental pollutants may play a relatively small role in the etiology of most of these conditions. For example, only about five to eight percent of congenital malformations are known to be caused by environmental exposures such as drugs, chemicals and ionizing radiation (Oakley, 1986). In the vast majority (65 to 70 percent) of cases of congenital anomalies, the cause is unknown. It is also important to recognize that many of these outcomes are interrelated

with one another, often in highly complex ways, and that these relationships may be influenced by the nature, timing, and intensity of exposure. For example, exposure to ionizing radiation during the period of organogenesis may result in fetal malformations, while exposures later in pregnancy have been linked to certain childhood malignancies. A low-level dose of radiation (e.g., < 5 rad) during the first trimester, may produce no apparent adverse effect; at a higher dose there may be evidence of fetal anomalies, while at extremely high levels spontaneous abortion may result.

A number of chemical and physical agents (2,4,5-T, dibromochloropropane, ionizing and nonionizing radiation, heat, noise) have been associated with deleterious reproductive effects in humans. Some of the major occupational and environmental agents associated with adverse reproductive effects are presented in Table 5.3.

Table 5.3. Environmental agents implicated in adverse reproductive outcomes

Exposure	Known or suspected effect
Anesthetic compounds	Infertility, spontaneous abortion, fetal malformations, low birth weight
Antineoplastic	Infertility, spontaneous abortion
Dibromochloropropane	Sperm abnormalities, infertility ionizing radiation infertility, microcephaly, chromosomal abnormalities, childhood malignancies
Lead	Infertility, spontaneous abortion, developmental disabilities
Manganese	Infertility
Organic mercury	Developmental disabilities, neurological abnormalities
Organic solvents	Congenital malformations, childhood malignancies
PCB's, PBB's	Fetal mortality, low birth weight, congenital abnormalities, developmental disabilities

Source: by permission: Aldrich & Griffith, 1993.

5.5 Cancer

A thorough discussion of carcinogenesis would be complex and extensive, as is the discussion of causation from an environmental perspective (Doll & Peto, 1981; Schottenfeld and Fraumeni, 1996). This discussion is only intended to acquaint the reader with a few key concepts involved in carcinogenesis, and to provide a foundation for additional reading.

It has been known for sometime that the process of carcinogenesis occurs at the cellular level (molecular). However, only through recent advances in molecular biology are we beginning to understand the process. Moolgavkar (1981) first proposed a two-step process for carcinogenesis, and coined the terms initiation and promotion. Initiation is the transformation of a benign or harmless cell to one with the potential for malignant growth. However, initiation is distinct from the event that causes a cell to grow rapidly and to acquire the capacity to metastasize.

Recently, the discovery that there is genetic material that may be passed from parent to offspring with the potential for causing cancer (oncogenes), and the genetic mapping of fragile sites on the genome, have enhanced general understanding of carcinogenesis. It is hypothesized that carcinogenesis requires a series of cellular events to occur, and that some of these events may require a specific sequence to be effective. For example, with familial breast cancer (Hall et al., 1990), and possibly with colon cancer (Kinzler et al., 1991), specific genes have been located on the genome that must be activated, as well as a subsequent loss of regulation for those genes. In both cases genetic changes favouring cell survival and growth are necessary adjuncts to the disease process, although the exact sequence of these sustaining events may not be as critical as with those involving the key genetic locus.

For the purposes of environmental epidemiology it is necessary to recognize the genetic-cellular nature of carcinogenesis and the interplay of multiple forces in order to achieve carcinogenesis. For example, although the role viruses play in carcinogenesis is not well understood, for several cancers (i.e., cervical cancer, liver cancer, nasopharyngeal cancer and Hodgkin's Disease) a role for a viral agent is well accepted.

Genetic factors

Clastogenesis, literally means chromosome damage. Chromosome damage takes the shape of deletions, breaks, gaps, rings, dicentrics, quadri radial figures, and acentric fragments. Gaps are defined as achromatic regions within a chromatid less than the width of the chromatid. Deletions in the nonbanded preparation are nonstaining regions in a chromatid greater than the width of the chromatid. Breaks are a discontinuity in a chromatid or chromosome that is misaligned. Rings, dicentrics, quadri radial figures, and acentric fragments result from exchanges within or between chromatids.

Similarly, mutational events, whether additions or deletions, is also a well accepted means for genetic transformation potentially relevant to cancer risk (initiation). Environmental agents may now be designated as mutagens, clastogens and mitogens (e.g., stimulating cell division). To these cell damaging ideas must be added the topics of cellular repair and regulation (Weinstein, 1991). Although cells are quite proficient at genetic repair, repair capabilities vary in human populations.

Detoxification

The ability to detoxify chemicals entering the body is highly variable between people. The liver acts as a first line defence for chemical hazards. The body prefers to remove or detoxify a hazardous chemical by combining it with glutathione in a process known as conjugation. However, there are limits to the body's store of glutathione and so alternative pathways for detoxification are required. Some of these are less efficient than others, so that errors may be made. Such errors may result in a chemical becoming more, rather than less toxic as it metabolizes through the system.

An example of such an error, can be shown in the activation of vinyl chloride as it metabolizes. When an alternative detoxification process to glutathione conjugation is involved with a vinyl chloride molecule, enzymes begin the process of breaking down the vinyl chloride molecule. However, because this process is inefficient, the breaking down process may be incomplete, and the partially metabolized vinyl chloride molecule remains, as a hyper-reactive epoxide (a much more dangerous agent). The epoxide seeks a chemical to stabilize its reactivity and will bind readily with DNA. This DNA bonding is a genetic level event that can produce a mutation. Vinyl chloride is considered a pro-carcinogen since it requires an action by the error-prone detoxification process to become a proximal carcinogen. There are similar terms for the extent to which an agent can both transform cells and promote cell growth. Agents may be termed initiators or promoters depending on their mode of action at the cellular level. Some agents have the capacity for both activities (e.g., cigarette smoke) so these may be termed complete carcinogen, while other agents are considered sufficient causes of cancer (i.e., they can produce carcinogenesis without, such as ionizing radiation).

Cancer is not simply one disease, but is commonly considered to be over 100 diseases. This perspective arises from the forty or so anatomic sites where cancer strikes and the various cell types that are present at many of these sites. Some cancers are more fully studied than others, and these are generally the more common cancers. Other cancers are considered to have relatively well established etiologies

Cancer epidemiology

Cancer is a historic disease, having been observed in ancient Egypt and Greece (Shimkin, 1977). Even dinosaurs skeletons have been discovered with cancers. However, cancer emerged as a major public health concern in the middle of this century. The emergence of cancer as a leading cause of death is closely linked to the eradication of infectious diseases as major causes of death and the extension of life expectancy. Heart disease, cancer and stroke are all diseases of aging, and as life expectancy rose, these causes of death rose as well, by attrition of the acute causes of death (Bailar and Smith, 1986).

Cancer rates have varied over this century (Figures 5.3 and 6.3). The most telling secular trend is the rise in lung cancer death rates and the declines of uterine, liver and stomach cancer mortality. Since 1960, the four major cancer sites have been lung, breast, colon-rectum and prostate. Lung cancer has come to have such an overwhelming impact that scientific publications have begun to report patterns separate from other cancers (SEER, 1989; Davis et al., 1990).

The pattern of cancer today in developing countries is very similar to the pattern shown for 1930-35 in developed countries (i.e., liver, stomach and uterine cancer predominate). As countries become more urbanized, industrialized, and westernized, cancer patterns shift to reflect those

found in the more developed countries. These changes, as well as studies of cancer patterns among migrants, are some of the strongest evidence for the assertion that most cancers arise from environmental sources. However, environment involves lifestyle and diet, as well as the ambient surroundings.

Following is a brief discussion of the general epidemiology of several cancers:

Lung Cancer. Lung cancer is the leading cause of cancer death in men and women. Increased risk is strongly associated with cigarette smoking -either direct or passive exposure. Occupational exposures include arsenic, organic chemicals, asbestos and ionizing radiation exposures (e.g., radon gas). Occupational risks are increased for smokers. Vitamin A deficiency is also a suspected risk factor.

Prostate Cancer. The highest occurrence of prostate cancer is among blacks, and is primarily a disease of older men (> 65 years). Familial association and risk from dietary fat are suspected. Cadmium is a potential occupational risk.

Breast Cancer. Breast cancer is the most common cancer in women. Although it is found in younger women, breast cancer is primarily a disease of post-menopausal women (e.g., > age 50). Increased risk is associated with family history, and with not bearing children prior to age 30. A role for dietary fat is suspected.

Colon-Rectal Cancer. Colo-rectal cancer is the third leading cause of cancer death in both men and women. It is associated with low fiber and/or high animal fat diet, and with a history of polyps and inflammatory bowel disease.

Cervical Cancer. Cervical cancer is a disease of young women, associated with early intercourse, multiple partners and cigarette smoking. Endometrial cancer is a disease of older women, associated with infertility, estrogen therapy and obesity.

Oral Cancer. Oral cancer is more common in men, and is associated with cigarette, cigar and pipe smoking, smokeless tobacco use and alcohol consumption.

Bladder Cancer. Bladder cancer is more common in whites, and in men. Smoking is recognized risk factor, and workers in dye, leather and rubber occupations are at higher risk.

Pancreatic Cancer. Pancreatic cancer is more common in blacks. Higher rates for males, and persons over age 65. Smoking is a recognized risk factor; dietary fat, chronic infections, diabetes and cirrhosis are suspected.

Skin Cancer. Common skin cancers occur in half of all people; however, a 98% survival rate has led to these cancers being excluded from most statistical reports. Melanoma is the most common lethal form. Fair complexion is a strong risk factor, as is excessive exposure to the sun. Coal tar, pitch or creosote, arsenic and radium are occupational risks.

Leukemia. Leukemia is a disease of children and older adults; both sexes, all races. Certain genetic risks are known (e.g., Down's Syndrome), as are viral agents (HTLV-1). Occupational risks are ionizing radiation and benzene.

Ovarian Cancer. Ovarian cancer is a disease of older women (> 60 years), and ones who have never borne children. Risk is increased by a history of breast, colo-rectal and endometrial cancer.

Brain Cancer. The increasing occurrence in recent years may be associated with job-related aromatic hydrocarbon exposures and non-ionizing radiation.

Lymphoma. Lymphoma has been increasing in recent years. This is essentially a mixed group of cancers with many suspected risk factors including agricultural chemicals, viruses and childhood exposures.

Stomach Cancer. It is a leading cancer in developing countries, associated with nitrates in food.

Liver Cancer. It is a leading cancer in developing countries, associated with Hepatitis B infection, cirrhosis and occupational exposure to aromatic hydrocarbons.

Investigating environmentally-related cancers

In studying environmental agents, there are many considerations, but among the most important is understanding the recognized risk factors. Much of the emphasis on environmental cancer epidemiology is directed to ecological studies. In this instance, the usefulness of standardized rates (age, race, and sex adjustment of disease rates) and expected numbers is well established, but with politically visible cancer, the emotional impact to a community may be accentuated even when expected disease patterns are observed. At any rate, it is important to know the expected risk in a community study, since this knowledge will be a valuable asset in setting study priorities and in educating the public. Similarly, knowing which cancers have established risk factors, and being able to evaluate the distribution of the risk factors in a community is also useful. Environmental epidemiologists should always review the cancer literature before undertaking a new study. However, the literature is so plentiful, with new hypotheses arising so quickly, it remains difficult to keep abreast of the most recent events.

Induction and latency will have profound influence on any environmentally related cancer investigation. In regard to cancer, the time to cell transformation and immortalization of the transformed cells is an example of the induction period of cancer. For chemically induced cancer in humans, the usual time frame exceeds 20 or more years from the time of exposure. With most studies, it is not unusual to assume an induction period of from 5 to 10 years. For most cancers, latency is believed to be 10 to 20 years. Time intervals are frequently included with case eligibility consideration to accommodate a latency interval (i.e., in considering an exposure one must look back in time for as much as 30 years).

With sparse populations and rare events, statistical artifacts become more of a real as opposed to potential problem. Thus, careful biologic reasoning is needed, complemented by subtle statistical methods. Epidemiologists should also look for consistency among the cases with the more acute events (e.g., birth defects, neurotoxicity), and with evidence from biologic markers of exposure or disease.

Chapter 6

Study Design and Methods

Learning objectives

- Appreciate the need for explicit hypotheses and clear objectives as a basis for study design.
- Be able to describe the major types of study design and their principal methods.
- Understand the strengths and limitations of each study design for environmental epidemiology.
- Understand the issues of bias, confounding, and statistical precision as they apply to designing environmental epidemiology studies.
- Be able to select an appropriate study design for addressing a specific environmental health problem.

6.1 Concepts of study design

The objective of an environmental epidemiology study is to assess whether an environmental exposure has an effect on health. This is achieved by comparing the health outcome occurrence of individuals who have been subject to different exposure levels and, in particular, by comparing the health outcomes of exposed persons with those of unexposed or lesser-exposed persons. The result of such a comparison is quantified by calculating a measure of effect, such as the ratio of the incidence rate among the exposed to that of the unexposed.

Sometimes the environmental hazard that caused the health effect is only vaguely suspected; in other situations it is totally unknown. The first step in investigating the health effect could then be a **descriptive study**, which is a study that defines the population group of interest, estimates the incidence or prevalence of the disease, and identifies any potential environmental hazards that might have caused the disease. A descriptive study does not analyse the statistical association between exposure and disease, but can be useful in creating hypotheses for further study.

If a specific cause–effect relationship is believed to exist, an **analytical study** can be carried out. In such studies a hypothesis about cause and effect is tested or a quantitative relationship between exposure and effect is evaluated. Analytical studies are based conceptually on following a population over a period of time. This population is called **source population** and its experience over time the **risk period** (person-time at risk). The concepts of populations used in this text are illustrated in Figure 3.1, page 42. Several study design options are available. The various study designs differ in the manner in which the **study population** is sampled from the source population, and the manner in which information is drawn from the risk period. Thus study designs may differ according to whether cases of disease are ascertained for a specified time period (incidence data), or at a particular point in time (prevalence data). Study designs may also differ according to whether they incorporate all of the information contained in the person-time

experience of the study population (cohort studies) or whether they attempt to obtain the same findings by comparing cases of disease with people without disease, selected as a sample of the person-time experience of the source population that generated the cases (case–control studies). Thus the difference between the analytical study types is not the direction of the association between exposure and health outcome, but the relative efficiency, feasibility and quality of data that can be obtained using the alternative approaches.

A third type of study design is the **experimental study**. In this type of study a defined population is divided into groups using a randomized sampling procedure and then the investigator will assign or administer an exposure (or treatment) to some of the groups. Analytical and descriptive studies have been called observational studies in order to distinguish them from experimental studies. Because deliberately exposing people to truly harmful agents is unethical, experimental studies are generally used only to measure the impact of treatments or preventive interventions.

Issues of timing

Studies are termed **prospective** when the investigator observes the population over time as events occur. Studies are termed **historical (retrospective)** when the investigator ascertains exposure or health outcomes after they have occurred — generally by using existing records. The issue of timing does not affect the study design, but it is important in relation to validity and feasibility. For example, a cohort study of a disease outcome with long induction period may cost less if existing records can identify a study population and classify exposure status for some time in the past. Additionally, the study would be more feasible since the investigator would not be obliged to wait through the entire induction period for the disease outcome events to occur. In general, studies based on existing data are less expensive than studies for which new data must be collected. However, the quality of existing data may not be adequate.

Hypotheses and aims

Studies are most likely to be productive if they are based on clearly stated hypotheses. A hypothesis can be developed from the results of earlier epidemiological studies or on the basis of biological or toxicological research. In most instances, the hypothesis is that an association exists between an environmental exposure and 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 and results show that there is an exposure–response relationship.

The specific aims to be accomplished should also be clearly understood and formulated. A statement of aims can clarify the issues to be addressed, the nature of the study population, and the principal exposure and outcome variables to be assessed.

Feasibility

The study's feasibility should be evaluated before the study is undertaken. Since most studies are observational rather than experimental, selection of an appropriate study population is the major criterion for study feasibility. Other considerations relate to the advantages and limitations regarding efficiency, logistics and susceptibility to bias, that pertain to the various study designs. Since it is commonly possible to address the issues using more than one study design, the investigator must consider each of these factors when selecting a study design.

6.2 Types of study

Table 6.1 shows the most commonly used types of epidemiological studies as listed in *Basic Epidemiology* (Beaglehole et al., 1993). All of these study designs have been used in environmental epidemiology.

Table 6.1. Types of epidemiological study design

Type of study	Alternative name	Unit of study
Descriptive studies		Individuals or populations
Analytical studies		
Ecological	Correlational	Populations
Cohort	Follow-up	Individuals
Historical cohort	Retrospective cohort	Individuals
Case–control	Case-reference	Individuals
Cross-sectional	Prevalence	Individuals
Experimental studies	*Intervention studies*	
Randomized controlled trial	Clinical trials	Patients
Field trials		Healthy people
Community trials		Communities

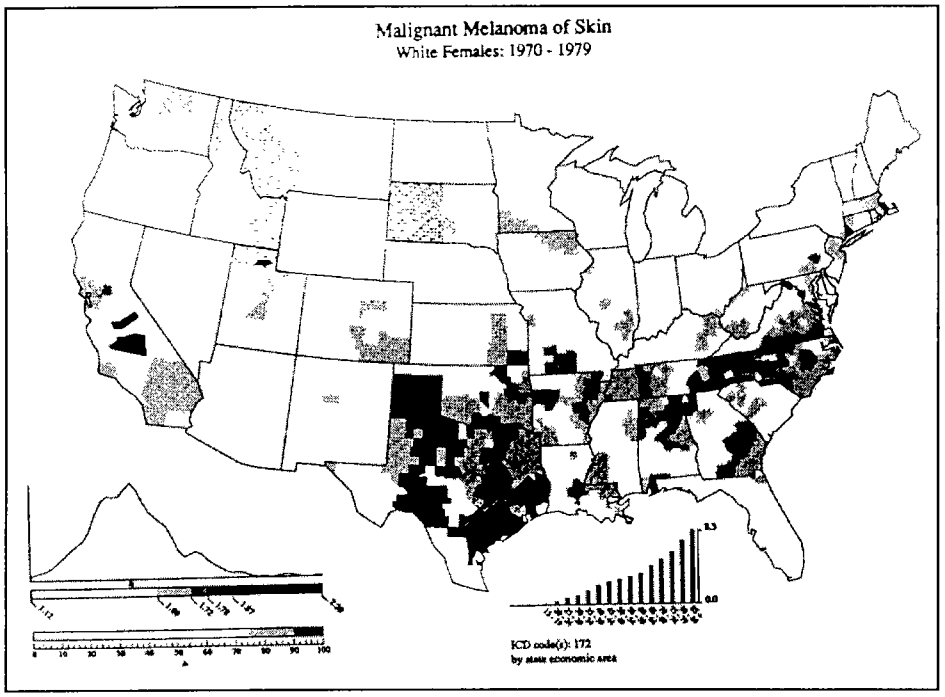
Source: adapted from Beaglehole et al., 1993.

Descriptive studies

Descriptive studies examine the distribution of disease and/or possible environmental determinants of disease in a defined population. They are often the first step in broader environmental health investigations because they may identify hypotheses to be evaluated in analytical studies. Descriptive studies do not formally evaluate the association between exposure and health outcome, although they can be helpful in assessing the possibility that an association exists. Furthermore, if a causal association is well-established, descriptive studies can be useful for establishing the presence or absence of an environmental health problem in a local population and quantifying its impact.

Most descriptive studies are based on existing mortality or morbidity statistics, such as hospital discharge data, and examine patterns of health outcome by age, gender or ethnicity, for specified time periods or geographical areas. For example, the cancer mortality map in Figure 6.1 depicts the incidence of malignant melanoma among white females in the United States. The darker colours indicate the population areas (census tracts) with the highest incidence of malignant melanoma. The gradient of cancer incidence rates with respect to latitude reflects the gradient in population exposure to sunlight (an association established scientifically by other studies).

Figure 6.1. Incidence of malignant melanoma among white females in the United States



Source: USEPA, 1987.

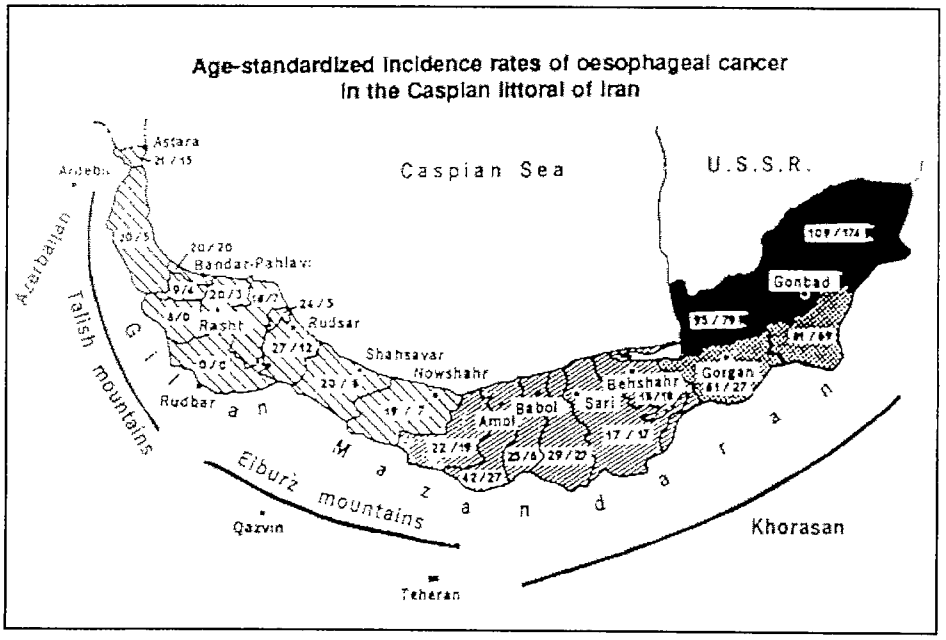
Descriptive data are commonly used to examine patterns of health outcome by place, time and person. **Geographical comparisons** based on standardized mortality and morbidity rates can be made among countries, or among regions within countries. Variations between countries with respect to rates of mortality attributable to cardiovascular disease and cancers have been the basis of hypotheses regarding the role of environmental factors in these diseases. However, international comparisons may be problematic due to differences in diagnostic practices and terminology. For example, the incidence rates of chronic bronchitis and emphysema formerly appeared to be considerably higher in some European countries than in the USA. But most of the apparent difference was in fact due to variations in diagnostic practice. In general, geographic contrasts between areas within a single country are likely to be less marked than those between countries, but can be more revealing in relation to environmental factors. An example is the nearly 30-fold difference in oesophageal cancer risk for women in different areas along the Caspian Sea in Iran (WHO, 1991b). Incidence rates (per 100 000 population) range from about 20 for males and 5 for females in some areas of the western provinces, to about 200 for males and 260 for females in parts of the eastern provinces (see Figure 6.2). Such large variations in incidence rate within a geographic area suggest the influence of environmental factors.

Temporal trends in mortality or morbidity rates can also be of value in indicating the possible effects of environmental factors. Figure 6.3 illustrates changes in male cancer mortality in the USA for the period 1930–1995, while Figure 5.3 shows female lung cancer mortality over time. The data for the males are consistent with increasing lung cancer incidence due to increased

exposure to tobacco smoke, and decreasing stomach cancer incidence due to greater availability of refrigeration and fresh foods.

Temporal trends can also be examined in order to evaluate the effectiveness of an intervention aimed at mitigating or eliminating harmful environmental exposures. The incidence of scurvy (due to lack of Vitamin C) declined dramatically among British sailors following the introduction of citrus fruit on trans-oceanic journeys. This provided evidence of an effective disease prevention strategy many years before the identification of Vitamin C in a laboratory.

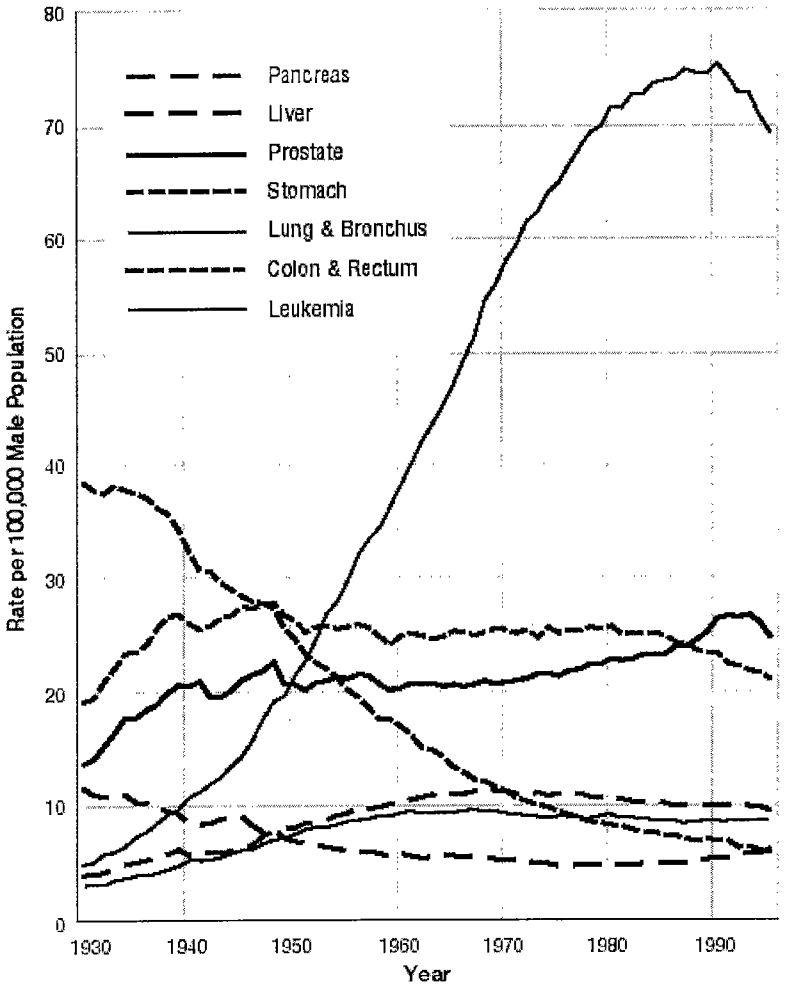
Figure 6.2. Oesophageal cancer among women in Iran



Source: WHO, 1991b.

Patterns of disease associated with **personal characteristics** may also provide insight about the effects of environmental factors. If, as is often the case, genetic factors are insufficient to fully explain associations observed between health outcomes and age, gender and ethnicity, environmental factors may be indicated. Sometimes, differences in exposure patterns between men and women can provide clues about contributory factors, as in the case of lung cancer. In the example concerning cancer mortality in the USA, death rates began to increase sooner in men (Figure 6.3) than in women (Figure 5.3); this was consistent with the earlier increase in smoking rates among males in most countries. More recently, in the USA and elsewhere, smoking rates have increased among females while decreasing among males. Patterns of tobacco-related cancer incidence and mortality should reflect these behaviours once the appropriate induction periods have been accounted for.

Figure 6.3. Secular change in male lung cancer mortality in USA, 1930-1995



Source: Landis et al, 1999; data from Vital Statistics, 1998.

Descriptive studies of migrant populations, in particular, can provide useful information about the relative roles of environmental and genetic factors (Parkin & Khlat, 1996). If records on country of origin are available, the health outcomes of migrants can be compared with those of their compatriots in both the country of origin and that of subsequent residence. For example, cancer morbidity and mortality rates in migrant populations often come to approximate those of the host country indicating that environmental factors play a role. If genetic factors were the exclusive cause, disease risks would not be influenced within one generation by migration.

For example, a descriptive study of cancer risk among migrants and their descendants in Israel found that risks for testis cancer, nasopharyngeal carcinoma and melanoma reflected the father's birthplace even in the second generation; while for ovarian, colorectal, cervical and thyroid cancers, differences in risk between the migrant groups had largely disappeared in the offspring (Parkin & Iscovich, 1997). The investigators concluded that environmental exposures were the

major causative factors for the latter group of cancers. There are many other examples of cancer studies based on migrants (eg Bouchardy et al, 1993; Hanley et al, 1995; Herrinton et al, 1994; Herrinton et al 1996; Rosenblatt et al, 1996).

Use of existing records

A substantial amount of data on environmental exposures and health outcome occurrence is collected routinely. Most countries have a system for recording and processing mortality data (WHO/CEC, 1989). A wide range of morbidity statistics, such as rates of congenital anomalies and infectious diseases, is also available in many countries. Additional information may be obtainable from periodic population surveys, such as the US National Health and Nutrition Survey. Sources of environmental exposure data are comparatively few. Moreover, the quality of much of this data is inadequate for epidemiological studies, since the data are not collected for the purpose of assessing health risk. Examples of environmental monitoring data include lists of industries containing estimates of the levels of pollutants discharged by geographic area; public health monitoring of water or commercial food products; and data collected during legally mandated monitoring of lead, pesticides, or other toxic substances in air or water.

Existing records can be used effectively to study the effects of past exposures and diseases of long latency. Many of the diseases that feature in environmental health studies, such as cancer, do not become manifest until many years after the biologically relevant exposure has occurred. Records of past exposure may therefore be required in order to evaluate current disease risk. This approach may be particularly appropriate if environmental exposure levels have changed over time. For example, recent exposures may be lower than in the past if newer, less polluting technologies and control techniques have been developed and applied. On the other hand, recent exposures may be higher due to increases in urbanization, industrialization and fuel-based transportation. In both instances, exposure assessment based on existing records of past exposure may provide a more accurate indication of the relevant exposure than newly-obtained data.

Use of existing records may also be worthwhile in terms of cost. Studies based on existing records tend to cost less and take less time than studies for which new data must be collected, and are useful as a prelude to more definitive but costly studies. In fact in most situations, the initial effort applied to examining existing records is justified because of the higher cost and length of time required to gather new data.

The principal disadvantages of using existing records are that the information they contain may not be appropriate to the study objectives and may not be sufficiently precise. For example, ambient air monitoring data routinely collected to identify when maximum allowable levels of air pollutants are exceeded may not provide specific enough information for estimating individual exposure levels. Other routinely-collected data such as cause of death on death certificates may be subject to substantial misclassification error (Kelsey et al., 1996). And although a number of countries now collect extensive data on morbidity, much care is required in using this type of information, even for descriptive studies.

However, strategies are being developed to improve the usefulness of existing information on environmental health hazards for decision-makers. For example, the *Health and Environment Analysis for Decision-making* (HEADLAMP) project coordinated by WHO has used traditional descriptive methods to demonstrate that routinely-collected environmental and health data can be linked and interpreted to provide decision-makers with sufficient information for devising

effective environmental health action plans (Briggs et al., 1996). The HEADLAMP process has three defining characteristics (Briggs, et al., 1996). First, the process focuses on already known relationships between environmental exposures and health effects. Environmental health indicators are identified on the basis of these relationships and are chosen for their potential value in policy decision-making. Second, the environmental health indicators are usually based on the use of routinely-collected data which, as discussed above, is cost-effective. Third, the aim of the HEADLAMP process is to generate information on which to pose preventive actions. Therefore, this process is oriented to environmental health management and prevention, rather than etiological research which is the focus of most epidemiological studies.

Ecological studies

Ecological studies are studies in which the investigators analyse hypothesized associations between environmental exposures and health outcomes using groups of people, rather than individuals as the unit of analysis. The numbers of individuals who have experienced specific combinations of exposure status and health outcome status are not known. Thus an ecological study compares aggregate measures of exposure, such as average exposure or proportion of population exposed, with aggregate measures of health outcome rates, for the same population.

The investigation of the association between arsenic and cancer mortality that was conducted in the endemic area of Blackfoot Disease in Taiwan is an example of an ecological study (Chen et al., 1985; Chen & Wang, 1990). The investigators used data on arsenic concentration measured in over 83 000 wells by the Taiwan Provincial Institute of Environmental Sanitation as a measure of arsenic exposure. These data were used to estimate the average arsenic concentrations in the water supplies of several hundred administrative districts. Cancer mortality data and population size data by administrative district were obtained from the Taiwan Provincial Department of Health. The investigators used correlation and regression analyses to examine the association between arsenic concentration in well-water and age-adjusted mortality for various malignant neoplasms, controlling for such factors as degree of urbanization and industrialization. Associations were observed between arsenic concentrations in well-water and elevated rates of cancers of the liver, nasal cavity, lung, skin, bladder and kidney.

Traditionally the role of ecological studies has been to perform exploratory analyses using existing population data. However, this study design plays a larger role in environmental epidemiology than in other areas of epidemiology because environmental factors often expose large populations in a similar manner. Thus air pollution in an urban area is likely to cause exposure to most inhabitants and their exposure might be uniformly different from that in a rural area. Similarly, the health effects following exposure to different concentrations of arsenic or products of chlorination in water supplies might be compared between areas. The potential for misclassification of individual-level exposure is less for these types of exposure. Thus an ecological study of environmental exposures can usually be interpreted with greater accuracy than studies of exposures that are determined by individual choice or behaviour, as in the case of cigarette smoking, or an exposure that has occurred via food.

Design of ecological studies

The strategy for conducting an ecological study is to determine whether those ecological units (i.e. population groups) with a high frequency of exposure also tend to be the groups with a high frequency of health outcome occurrence. Generally, the investigator obtains group rates of the health outcome and a measure of exposure prevalence for the same group. These data are then analysed statistically to estimate the group-level associations.

In common with the approaches used for descriptive studies, ecological analyses often use **geographical areas**, such as countries or administrative units within countries, as the basis for defining the groups. This was demonstrated by the aforementioned study of arsenic in well-water and cancer mortality (Chen & Wang, 1990). Similar studies have been conducted in other countries to evaluate the risk of cancer due to consumption of water containing chlorination products (Bean et al., 1982a), radioactive compounds (Bean et al., 1982b), and volatile organic chemicals (Fagliano et al., 1990). The ecological study design has also been used to compare rates of congenital malformations and low birthweight among US census tracts exposed to different levels of environmental contamination (Shaw et al., 1992) and cancer mortality rates among US counties with or without identified hazardous waste sites as potential exposure sources (Griffith et al., 1989).

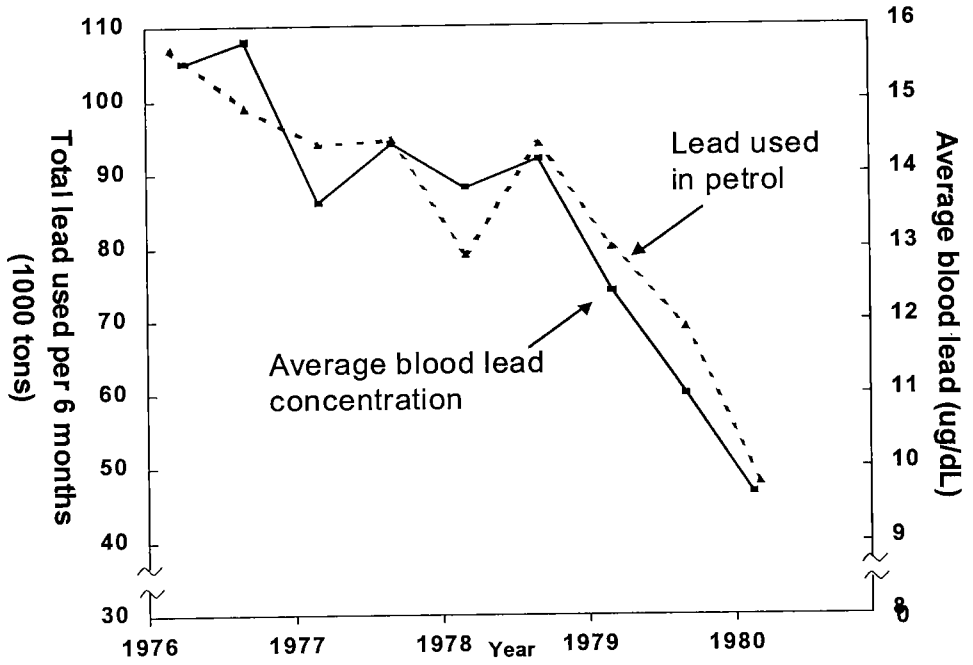
Although this approach is straightforward, determining exposure levels precisely can be difficult if the administrative units providing the health outcome data bear little relation to the occurrence or distribution of the environmental factor being studied. Water monitoring data, for instance, may not be collected specifically for an administrative unit such as a township; yet this may be the unit of analysis for which the health outcome rates are to be estimated. Estimating average exposure for an administrative unit on the basis of available monitoring data is therefore often necessary. Thus in a study of the association between air pollution and lung function, individuals in each census tract of a study were assigned the averages of pollution values obtained by existing air monitoring stations located within 10 miles of the population centroid of the census tract (Schwartz, 1989). Clearly, when planning an epidemiological analysis based on geographic areas, an evaluation must be made of how well the available data on exposures and health outcomes match the selected units of analysis.

Controlling for risk factors that may cause bias and confounding in ecological studies needs to be considered because groups of people from different geographic areas often live in different conditions and have different lifestyles. This means that the specific exposure under study, and exposure to other risk factors for the same disease, may vary between groups. Confounding or effect modification cannot be ruled out on the basis of other studies of individuals having demonstrated lack of confounding within each study area (Morgenstern & Thomas, 1993). For example, within one area of high air pollution there may be no association between smoking habits and air pollution exposure. The same may also be true of an area of low air pollution. Yet if average air pollution and average smoking habits are correlated, an ecological study of air pollution effects on the lung will suffer from confounding because some of the apparent association between the lung effects and air pollution may actually be due to different rates of cigarette smoking in the groups.

The groups studied in an ecological study may be defined by **time period**. Temporal comparisons or time trend studies examine associations between changes in exposure to environmental factors and changes in health outcome rates over time within the same base

population. One of the advantages of making a temporal comparison is that population characteristics, such as socio-demographic factors, that could bias group-level associations, may be relatively constant over time in the same geographical area. Figure 6.4 shows the correlation between the amount of lead used in petrol in the US and average blood lead concentrations as measured by an ecological study, for a four-year period (Goldsmith, 1988).

Figure 6.4. Changes in the amount of lead used in petrol and average blood lead concentrations in the United States



Source: Goldsmith, 1988.

Strengths and limitations

The principal strength of ecological studies is that they are usually based on existing data and therefore relatively inexpensive to conduct. This study design can be useful for studying rare diseases caused by relatively rare exposures. This is because the source population for the ecological comparisons can be very large, as when comparisons between countries are made.

However, ecological studies can be difficult to interpret since information about factors that could bias the findings is often limited. An important consideration is the potential for **ecological bias**, in which the group-level associations do not reflect individual-level associations accurately (Morgenstern, 1982; Greenland & Morgenstern, 1989; Morgenstern & Thomas, 1993). The underlying problem is that the use of aggregate measures of exposure may not reflect individual-level exposure status adequately due to within-group heterogeneity in exposure (Morgenstern & Thomas, 1993). This study design must therefore be used with caution because the magnitude of ecological bias is likely to be more severe and less predictable than individual bias in estimating the same effect.

Ecological studies have further limitations. For example, non-differential exposure misclassification can lead to an overestimation of the measure of effect; it should not be assumed that bias due to inadequate information would result in underestimation only of the exposure effect (Brenner et al., 1992). Moreover, environmental exposures are often highly correlated which means that identifying the causal exposure can be difficult. In a study of ambient air pollution, for instance, different regions may have highly correlated concentrations of ozone, acid aerosols and respirable particulates, making separation of the particular effect of each pollutant problematic. Finally, the analysis of ecological data may be limited if the study includes relatively few ecological units of observation, and even if the units contain very large populations. (The statistical precision of ecological studies is based primarily on the number of units and not on the size of the populations that constitute the units.)

The study of time trends over a long period of time may also be complicated by substantial changes in demographics and baseline risk characteristics. In recent decades such changes have occurred in many developing countries due to increasing population density and industrialization. Furthermore, data on exposure and health outcome rates may not be comparable. For example, techniques and instruments to measure air pollution have changed over time. Diagnostic practices and disease coding schemes such as the International Classification of Diseases have also changed. Investigators should be aware of such changes and develop strategies for rendering data sufficiently comparable.

Despite the potential shortcomings of ecological studies, however, they can demonstrate clearly that a change in exposure levels is followed by a change in the plausible health outcomes. This applies especially if a temporal analysis is combined with a geographical analysis to show that a change over time in exposure in one area is associated with a change in health outcome, while in otherwise comparable geographical areas, exposures and health outcome rates remain stable.

Cohort studies

In a **cohort study**, the study population consists of individuals who are at risk of developing a particular disease or health outcome. The individuals are divided into groups according to their exposure status. The groups are then followed over time to determine the subsequent incidence of the health outcome within each group. A cohort study therefore enables an investigator to measure incidence rates and to estimate all effect measures, such as rate ratios and rate differences, for multiple health outcomes. A cohort study is termed **prospective** if the data are collected as the events unfold and **historical** if it examines past events (using existing records); these terms refer to timing and not to study design.

Cohort studies have been used effectively in environmental epidemiology to assess the long-term health effects of acute exposure to environmental hazards. For instance, cohort studies have followed populations exposed to radiation, including: populations in Japan who experienced the nuclear bomb explosions during World War II; children who underwent radiation treatment for an enlarged thymus; women treated with radiation for post-partum mastitis, and communities who were affected by the Chernobyl nuclear accident. Cohort studies have also been undertaken to measure chronic effects in populations exposed to environmental releases of toxic chemicals from industrial facilities. The study of communities living in the vicinity of the pesticide factory in Bhopal, India, from which methyl-isocynate leakage killed more than 2000 people and

poisoned 200 000 others in 1984, is one example. The study of residents in Seveso, Italy, who were potentially exposed to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin following an explosion at a chemical manufacturing plant in 1976 is another (Bertazzi et al., 1989; Bertazzi et al., 1992). Additional classic examples of cohort studies following poisoning episodes include studies in Japan of populations who were exposed to cadmium and developed Itai-itai disease as a result (WHO, 1992a), studies of populations suffering from Yusho Disease due to ingestion of rice oil contaminated with polychlorinated biphenyls, and studies of children living near Minamata who had been exposed to methyl mercury (WHO, 1990, 1992c).

Design of cohort studies

There are two basic types of cohort. Each is defined according to the type of source population upon which it is based. A **fixed cohort** consists of a population of individuals who are identified at a point or interval of time and then followed over time. Fixed cohorts can also be defined on the basis of a specific event such as exposure to a chemical spill. The cohort is considered fixed because cohort membership is defined (i.e. fixed) at the time of entry. Examples include the children irradiated for an enlarged thymus, or the communities in Seveso, Italy, who were potentially exposed to dioxin. A **dynamic cohort** allows the inclusion of members over time as they fit the selection criteria. An example of a dynamic cohort would be a community, exposed to air pollution, for instance into and out of which individuals could move during the follow-up period. The concept of a dynamic cohort is attractive because it reflects the constantly changing nature of populations. Dynamic cohort studies are most feasible in countries or regions in which population identification or registration records are routinely maintained.

The principal steps in conducting a cohort study are:

- identify and select the study population;
- classify the study population according to exposure status and other relevant risk factors;
- follow the cohort members over time to determine the health outcome occurrence by exposure status sub-groups.

Selection of study population

Cohort members must be **at risk** of developing the health outcome of interest. Therefore, when selecting the cohort the investigator must eliminate individuals who cannot develop the health outcome, and individuals who already have the outcome of interest, as well as persons who are not susceptible (See Box 6.1). For example, in a study of persons exposed to environmental tobacco smoke, potential cohort members may have to be screened at the beginning of the study to ascertain that they are free of respiratory disease.

The study population (cohort) can be selected in one of two ways. Subjects can be recruited into the cohort without regard to exposure status — i.e. the **census** strategy. Exposure status is determined after subjects have been enrolled into the cohort. This approach is useful if the exposure is likely to be common or if the aim of the study is to identify effects of more than one environmental factor since "exposure" can be classified in different ways for the same cohort. For example, in a cohort study of several environmental factors associated with the onset of wheezing among men, investigators identified a cohort of community-dwelling middle-aged and older men who had no history of wheezing or asthma, without regard to their exposure histories (Sparrow et al., 1993). The investigators then obtained information about each of the different exposures and performed a variety of medical tests at a baseline examination. The analysis

examined which of the exposures and baseline characteristics were predictors of new onset wheezing for the planned three–year follow-up period. But census sampling for a cohort study can be inefficient if the exposure of interest is rare.

Box 6.1. Guidelines for selecting exposed and unexposed subjects in cohort studies

- Unexposed persons should be sampled from the same (or comparable) population as the exposed group.
- Both exposed and unexposed groups should be free of the health outcome of interest and equally susceptible to development of the outcome at the beginning of the study.
- The baseline characteristics of exposed persons should not differ systematically from those of unexposed persons except regarding the exposure of interest.
- Equivalent quality and quantity of information should be available on exposure and health outcome status for the exposed and unexposed groups.
- Both groups should be accessible and available for follow-up.
- Multiple comparison groups of unexposed subjects chosen on the basis of different criteria may reinforce the validity of the findings.

Source: adapted from Greenberg & Daniels, 1993.

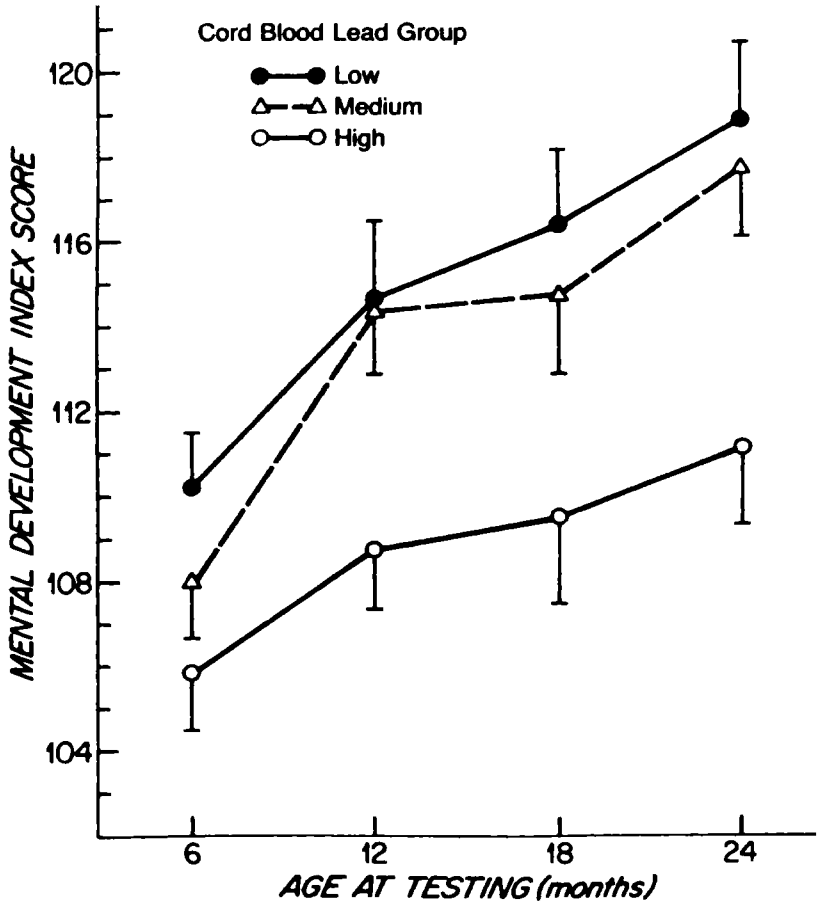
The alternative strategy is to select members for the cohort on the basis of their **exposure status**. This approach can be efficient since balanced numbers of exposed and unexposed subjects can be selected. For example, in a study of prenatal lead exposure and mental development, umbilical cord blood lead concentrations were measured for over 11 000 newborn infants. The findings from this survey were used to identify those children whose blood lead levels fell below the 10th percentile (< 3 μg/dL), near the 50th percentile (6 to 7 μg/dL), and above the 90th percentile (≥10 μg/dL) (Bellinger et al., 1987). These groups representing low, medium, and high lead-exposed children, respectively, were then followed to 24 months of age using standardized tests of cognitive development. As shown in Figure 6.5, infants in the high-lead prenatal exposure group scored less than infants in the other two groups, throughout the follow-up period. The study was thus able to demonstrate the adverse effect of prenatal and postnatal lead exposure on the mental development of children, but without having to follow the entire source population of 11 000 newborns.

When planning a cohort study, identification of an appropriate **sampling frame** for selecting the study population is crucial (see Chapter 3, pp. 61, for discussion of sampling). Sampling frames for cohort studies include population census lists, church parish registries, driver's licence lists, and telephone directories. Evidently, the availability of useful sampling frames varies considerably among communities, and in some areas, may not be available at all.

Well-defined study populations are often favoured by environmental epidemiologists since, provided they are appropriate for addressing the study hypotheses, they are often relatively easy to identify and follow. Examples include occupational groups (chosen not necessarily because of a particular occupational exposure, but on account of their fixed location, or the availability of records pertaining to them); members of prepaid health plans; and schools, college or military populations. The investigator must be certain, however, that the advantages of ease of

identification and follow-up are not cancelled out by constraints on interpretation; for example, the study population could be so atypical that the study findings might not be generaliseable.

Figure 6.5. Mean mental development index score by cord blood lead concentration



Source: Bellinger et al., 1987.

Selection of a comparison group

In cohort studies, a non-exposed group is the comparison or reference group. The comparison group provides a measure of the health outcomes that occur in a population without exposure. As mentioned above, this group could emerge from the defined cohort (after it has been divided into exposed and unexposed groups), in which case it is known as an **internal** comparison group. It is also possible to choose a comparison group from a broader sample of the source population which is known to be unexposed. In this case, the comparison group is an **external** comparison group. Some studies have both an external and an internal comparison group. Thus in a study of health effects among residents living near a hazardous waste disposal site, investigators selected residents in a nearby non-exposed community as an external comparison group, and also divided residents within the disposal site community into groups based on estimated potential exposure, including one that was non-exposed (Baker et al., 1988).

The advantage of using an internal comparison group is that the various exposure groups within the same study population are more likely to be comparable than using an external comparison population. Yet identifying a large enough non-exposed group within a single cohort may not be possible. Moreover, by using an external comparison group, the investigator may be able to use existing data, rather than collecting new data, thus making considerable cost savings. For example, a study could compare the number of deaths in an exposed community with the number of deaths — based on regional or national mortality rates derived from existing records — that would be expected under normal circumstances. The major challenge when selecting an external comparison group is to ensure that, apart from being non-exposed, it is comparable to the exposed population in every respect.

Population follow-up

Methods for following cohorts depend on the availability of adequate records. So a check should be made when designing a cohort study to ensure that the requisite records exist. If mortality records are to be used, the investigator must decide what sources of records to include in the study, e.g. national death registries or death certificates. Obtaining mortality records on a national or international basis may be too costly. On the other hand, limited records might not provide accurate or sufficient information on subjects who migrate during the follow-up period.

Follow-up studies of non-fatal health outcomes are no different in principle from mortality cohort studies, but since measurement of non-fatal health outcomes usually necessitates contact with the subjects, more active methods are likely to be required to follow the study population. Direct contact would be necessary, for instance, to obtain information on outcomes such as respiratory symptoms or to measure lung function. That said, direct contact would be unnecessary if the health outcome could be determined from existing data sources, as in the case of hospitalization for a myocardial infarction. Typically, in long-term cohort studies, investigators contact the subjects periodically. So in designing a study, the investigator must strike a balance between incurring higher costs (as a result of more frequent follow-up intervals), and sustaining greater losses of information (as a result of less frequent subject contact).

Historical cohort studies

A historical cohort study is conceptually identical to a prospective cohort study except that the study takes place after the causal events have unfolded. The investigator typically uses existing records to identify a cohort for some time in the past, to assign exposure status, and then to follow the cohort forward over time. For example, in the cohort studies of children exposed to ionizing radiation during treatment of an enlarged thymus, the investigators used medical records to identify the children and to estimate their radiation dose. Cancer registry records and vital records enabled investigators to follow the children to determine cancer incidence and mortality.

The relative advantages of prospective and historical cohort studies are summarized in Table 6.2. Historical cohort studies are generally less expensive and take less time to complete than prospective cohort studies. However, existing records upon which historical cohort studies are based, may not provide accurate information. And even if information on the environmental factor of interest is available, information on confounding risk factors may not be. Moreover, exposure monitoring methods and diagnostic criteria may have changed, making it difficult to combine data that was collected some time ago with data that was collected more recently.

Table 6.2. Comparison of prospective and historical approach for cohort studies

Attribute	Prospective Approach	Historical Approach
Information	More complete and accurate	Less complete and accurate
Discontinued exposure	Not useful	Useful
Emerging, new exposures	Useful	Not useful
Expense	More costly	Less costly
Completion time	Longer	Shorter

Source: adapted from Greenberg & Daniels, 1993.

Strengths and limitations

The cohort study design is generally regarded as the most definitive of the observational study designs because the investigator identifies the study population and then follows the causal events as they unfold: from exposure until the development of the health outcome. Exposure status is determined before the health outcome events occur. Therefore, knowledge of health outcome risk cannot influence how exposure is classified. The temporal relationship between exposure and health outcome is clearly determined. This study design also allows for the analysis of multiple health outcomes in relation to an exposure.

However, cohort studies are not commonly undertaken in environmental epidemiology since they can be expensive. And they are likely to be infeasible if the latency period between exposure and the health outcome occurrence is lengthy. Furthermore, cohort studies can be inefficient if the health outcome is rare. This is because the study population would have to be very large in order for a sufficient number of outcome events to occur. Finally, in comparison to studies of occupational cohorts, identifying and following cohorts in a general population is more difficult.

Cohort studies have been used effectively to evaluate health outcomes that do not have long latency periods (i.e. within months to several years). An example is the study presented of children followed from birth until 24 months of age to assess the effect of pre-natal and post-natal lead exposure on cognitive development (Bellinger et al., 1987). Another example is the study by Neas et al. (1994) of the effects of exposure to indoor particulate matter on respiratory symptoms and lung function among children over a five–year period. The follow-up period of a cohort study may be of any length, provided that an adequate number of outcome events (estimated on the basis of the size of the study population, the expected incidence rates, and the induction period) is likely to occur.

Although long-term cohort studies can be expensive and difficult to conduct, they may be justified if evidence strongly suggests that a chronic health effect is associated with an environmental exposure. Needleman et al. (1990) conducted an 11–year cohort study of children exposed to lead, as measured by dentine lead levels, to assess the presumed association of lead exposure with long-term adverse neurobehavioural functioning. The decision to undertake this

long-term study was based on findings of cross-sectional and short-term cohort studies which had revealed an apparent association between lead exposure and adverse health effects in children. Similarly, Dockery et al. (1993b) evaluated the effects of air pollution on mortality for a 14 to 16 year follow-up period, and in so doing demonstrated associations between fine-particulate air pollution and mortality due to lung cancer and cardiopulmonary disease. Their study was prompted by ecological and time-series studies that had previously indicated such an association on a cross-sectional basis. As is often the case with long-term epidemiology research, the study populations in these studies came to be regarded as a valuable scientific resource.

Case–control studies

A case–control study (or case-reference study) examines the association between exposure and a health outcome by comparing **cases**, or individuals who develop the outcome, and **controls** who are a sample of the source population from which the cases were identified. As discussed in Chapter 3, the source population and person-time experience of the source population are not directly observed in a case-control study. Controls are usually individuals who are similar to the cases in terms of risk characteristics, but who have not developed the health outcome. Having selected cases and controls, the investigator then determines the prior exposure status of the cases and controls by examining existing records and questionnaire information or measuring biomarkers of past exposure. The distribution of exposure for the controls provides an indication of the exposure distribution in the source population, and would also be the "expected" exposure distribution for the cases if no association existed between exposure and the health outcome.

Use of the case–control study design has increased since investigators have come to appreciate its efficiency and its conceptual relationship to the cohort study design (Breslow & Day 1980; Rothman & Greenland, 1998). The efficiency of the case-control study derives from not having to wait for disease occurrence during the induction period of a chronic disease and from limiting the expense of exposure measurement to cases and controls (i.e. exposure is not measured for the entire source population). Examples of case–control studies include outbreak investigations such as the investigation of the asthma epidemic in Barcelona, Spain (Antó et al., 1989; Antó & Sunyer, 1990) and of Toxic Oil Syndrome (TOS), also in Spain. In the latter study, the investigators obtained the food ingestion histories of 124 cases with TOS and of 124 individuals without the disease. Both groups had similar socioeconomic backgrounds. All patients with TOS reported consumption of illegally marketed cooking oil, whereas only 6.4%of the control group had consumed the oil. This observation led to further studies which established the role of the contaminated oil in causing TOS, even though the actual toxin in the oil was not identified.

Case–control studies have been used extensively to study the role of environmental factors in the development of various types of cancer. Examples include studies of lung cancer and residential exposure to radon (Samet, 1989; Lubin, 1994); childhood tumours and leukaemia and residential exposure to electromagnetic radiation (London et al., 1991; Savitz et al., 1993; Washburn et al., 1994); breast cancer and exposure to pesticides such as *p,p'*-DDT (Wolff et al., 1993; Krieger et al., 1994); lung cancer and indoor air pollution in China (Liu et al., 1993); and leukaemia in individuals exposed to radioactive fallout from nuclear testing facilities in the USA (Stevens et al., 1990). This study design has also been used to study birth defects and other reproductive outcomes associated with exposure to pesticides and drinking-water contaminants (see, for example, Goldberg et al., 1990).

The case-control design has also been used to increase the efficiency of studying cohort populations. If sufficient information is available for identifying a cohort, but the expense of obtaining detailed information on exposure status or confounding factors is prohibitive, the investigator may be able to conduct a case–control analysis within a cohort study population. If so, data on exposure status and other risk factors will be required only for the cases and controls. This design is referred to as a **nested case–control** study.

Design of case–control studies

The principal steps in conducting a case–control study are:

- establish a case definition;
- establish criteria for a study population and identify a sampling frame for selecting cases;
- identify an appropriate sampling frame for selecting controls;
- assess prior exposure status and other relevant characteristics of the cases and controls;
- compare the prior exposure experiences of cases and controls in order to estimate the association between the exposures and health outcome.

Case identification

Generally case–control studies are based on health outcomes that are recorded officially, such as mortality or cancer incidence in areas for which tumour registries exist, or conditions requiring hospital admission. Use of such records is cost-efficient. Examples include the studies of lung cancer, childhood cancer, leukaemia, and birth defects mentioned previously.

Many case–control studies are based on patients admitted to hospitals or seen for diseases at medical clinics. Such studies have several advantages: the subjects and their records are highly accessible; the subjects are generally cooperative; the background characteristics of cases and controls can be balanced if both attend the same facility, and the availability of medical records and biological specimens makes collection of exposure information relatively easy. But using hospital-based cases is appropriate only if the health outcomes that are being investigated usually result in hospitalization. So although selecting hospitalized cases may be suitable for studying conditions such as cancer, this would not be so for studying dermatitis. If hospital patients are to be selected, the investigator should understand the referral patterns to the hospital so as to be able to identify the true base population.

Population-based cases are the principal alternative to hospital-based cases. If a population-based approach is adopted, this means that all incident cases of the outcome in a defined geographical area are included as cases, irrespective of the source or mechanism by which they were identified. A combination of case reporting by all hospitals in a region, as well as other health care treatment facilities and offices of medical practitioners, might therefore be used. The advantages of a population-based case approach are that: the source population is better defined than for hospital-based cases; ascertaining that cases and controls derive from the same source population is fairly straightforward, and the exposure histories of the cases are more likely to reflect those of persons in the source population who do not have the disease of interest. The disadvantage of this approach is that high costs may be incurred when establishing the network of personnel required for identifying cases from multiple sources.

Selection of controls

The choice of a control group is dictated by:

- the source of the cases;
- the relative costs of different approaches to identify controls;
- the facilities available to the investigator carrying out the study.

Controls should be individuals from the source population who would have been considered cases if they had developed the health outcome (Axelson, 1985). The three most commonly used control groups are: a random sample of the population from which the cases were selected; persons seeking medical care at the same institutions as the cases for conditions believed to be unrelated to the health outcome of interest; and neighbours of the cases. Less frequently used controls include friends, schoolmates, siblings and fellow workers.

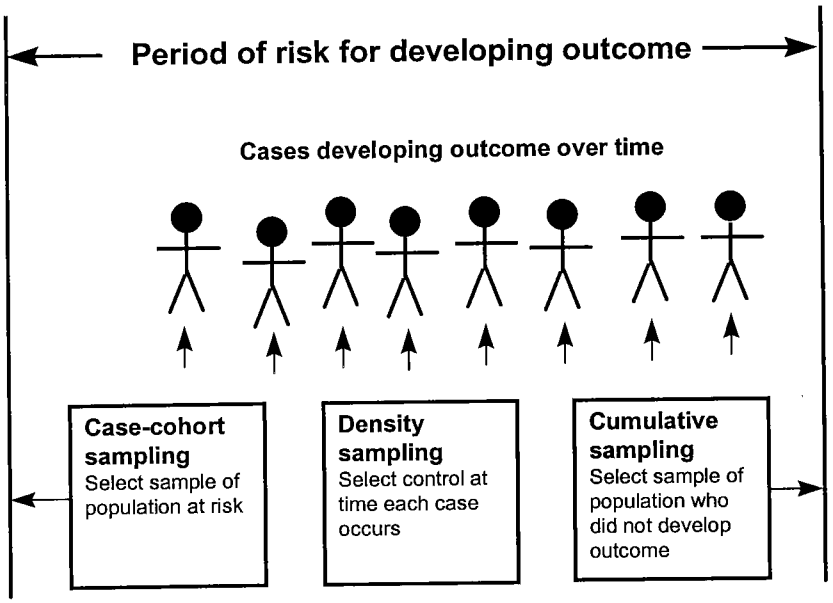
Although controls are individuals that could develop the health outcome of interest, they do not have to be disease–free. Indeed, using persons with other diseases as controls may be efficient. For example, in a study of lung cancer using a population-based tumour registry, defining controls as those individuals who have been registered as having developed cancer at sites unrelated to the environmental exposures of interest, may be efficient. Similarly, if cases are identified through hospital records, other hospital patients can often be used as controls. For example, Liu and colleagues studied the role of indoor air pollution and lung cancer in China by selecting incident lung cancer cases and individually matched controls from among patients admitted to the surgical departments of the same hospital (Liu et al., 1993). Patients who had been admitted for malignant tumours, chronic obstructive lung diseases, or coronary heart disease were excluded to minimize confounding due to cigarette smoking.

If cases consist of all those individuals within a defined population who develop the disease, the best control group would generally be a sample of individuals from the same source population. But identifying the control group can be expensive unless the sampling can be based on existing records, such as national security identifications, drivers' licenses, or telephone directories. Methods have been developed, such as random digit dialling, so that the telephone can be used to identify control groups that are representative of a large base population.

Another important consideration in identifying sources of controls is the type of information to be obtained. For example, if tissue or any other type of body specimen must be obtained from the control to verify the diagnosis, or to measure a biological marker of exposure or susceptibility, controls may have to be selected from among hospitalized patients or persons attending medical clinics. In selecting a control population it may be necessary to balance, for example, the easier logistics of using hospital controls, against the greater generalisability that use of population-based controls would afford.

There are three major approaches to selecting controls — cumulative sampling, density sampling, and case-cohort sampling — depending on when the controls are selected in relation to the incident cases. (See Figure 6.6) **Cumulative sampling** involves selecting controls from those people in the source population who have not developed the health outcome by the end of the period of observation. This approach is popular because it is straightforward, but it renders interpretation difficult for a dynamic population since the periods of observation are not necessarily the same for the individuals who are being compared (since the controls survived until the end of the observation period while the cases did not).

Figure 6.6. Sampling of controls in a case-control study



Density sampling (risk-set sampling) is the most common approach for population-based case–control studies. Controls are selected from persons in the population who are "at risk" (i.e., those who could develop the health outcome) at the time of onset of each case. In other words, whenever a case occurs, the investigator samples one or more controls who have not yet developed the health outcome from the source population. The control group therefore provides an accurate representation of the risk period when the cases actually develop. A case–control study of lung cancer and radon exposure, for which random-digit telephone dialling is used to identify a population-based control for every case reported to a tumour registry on an ongoing basis, would be an example of density sampling.

Case-cohort sampling is not as widely used as either cumulative incidence sampling or density sampling. A control group is sampled from the source population at the beginning of the period of risk, before any cases have occurred. An example might be a study of the residents of an area that has been selected as the location for a planned industrial facility. Prior to constructing the facility, a random sample of the community residents could be made. Subsequently, the residential location of incident cases would be compared with the residential locations of the controls selected at the beginning of the study. Thus an estimate of the distribution of exposure in the base population is calculated before the onset of any cases; it is also often possible to use the same control group to examine different health outcomes. But this study design can be used only if an entire source population, or cohort, can be clearly identified. It accordingly plays a greater role in occupational epidemiology than in environmental epidemiology studies since occupational cohorts are fairly easy to identify provided employment records are available.

Matching

Matching is a strategy for selecting controls so that the distributions of some *a priori* selected risk factors are identical or nearly the same for the controls as for the cases. Efficiency of the statistical analysis is enhanced because the cases and controls are as comparable as possible in relation to other variables that may confound the association.

In a study of radon exposure and lung cancer, controls could be matched to cases for age, gender and history of cigarette smoking, since each of these could be a risk factor for lung cancer and could also be associated with cumulative exposure to radon. If matching was not performed for such a study, the cases would probably include a much larger proportion of older persons, males, and persons who smoked cigarettes than would the controls randomly selected from the population. Evaluation of the association between radon exposure and lung cancer incidence, while analytically adjusting for these other factors would be inefficient since it would necessitate use of a larger control group.

With the exception of very small study populations or unique distributions of risk factors among cases, investigators have generally found frequency matching or probability-based sampling approaches to enhance comparability of the case and control study groups, to be adequate. (See Section 6.3 for further discussion of matching.)

Strengths and limitations

The case–control study design has several attractive features. It is efficient for studying rare diseases, especially those with a long induction period, and is often less expensive and less time-consuming to apply than a prospective cohort study design. Furthermore, by selecting study subjects on the basis of a health outcome, the investigator can examine the role of multiple environmental factors.

Case–control studies are also effective for studying dynamic populations which could be difficult to follow if a cohort study design was used. If, for example, an association between an environmental exposure, such as radon, and the incidence of lung cancer was being evaluated, it would be difficult to keep track of those members of the cohort who moved in or out of the exposed community. As an alternative, a sample of the dynamic population consisting of the individuals who had developed lung cancer, and another sample of similar individuals from the same population who had not developed lung cancer at the time the cases were diagnosed, could provide information for estimating the distribution of exposure in the dynamic population.

Two of the principal disadvantages of the case–control study design are that identifying an appropriate control population can be difficult, and there can be greater bias in classifying exposure status since this is determined after the outcome has developed. This bias may occur in one of several ways. The health outcome itself may affect the exposure measure directly or indirectly. For example, individuals with respiratory symptoms might deliberately avoid exposure to tobacco smoke. If they subsequently participated in a case–control study of respiratory disease that study would probably underestimate the association between respiratory disease and exposure to cigarette smoke. Such bias may also be of concern if biological markers are being relied on as an indicator of past exposure. For example, for cancer cases, some investigators have measured organochlorine pesticide concentrations in adipose tissue as an indicator of prior exposure. But this measure may not have reflected past exposure accurately if the onset of the disease was associated with metabolic changes that alter body storage and

distribution of the pesticides (Wolff et al., 1993). Another problem area concerns individuals' recall of exposure. Many case–control studies rely on subjects to report on earlier exposures, but cases tend to recall exposures more so than controls who have less motivation to recall earlier events. It is also possible that the study staff's awareness of the subjects' disease status may affect their measurement or recording of exposure status.

Another disadvantage of the case–control study design is that the analysis cannot yield estimates of health outcome incidence rates, rate differences, or attributable risk directly, although estimating these parameters in population-based studies is possible.

Cross-sectional studies

Cross-sectional studies examine the association between an environmental exposure and disease prevalence, i.e. the proportion of the population affected by the health outcome at a particular point in time or during a short period of time. Estimates of exposures and measurements of personal characteristics and biological effects are made at the same time. For example, investigators may study lung function in an apparently healthy study population and at the same time ask questions about exposure to indoor air pollution and environmental tobacco smoke. Table 3.2 on page 50 presents the findings of a cross-sectional study of respiratory symptoms prevalence and exposure to automobile exhaust, as an example.

Cross-sectional studies are usually undertaken before cohort or case–control studies because they tend to cost less and take less time to complete. However, cross-sectional studies are not merely quick forms of the latter types of studies since they examine prevalent rather than incident events. Thus a cross-sectional study design might be used to evaluate the risk of chronic lung disease due to air pollution, but it could not be used to examine respiratory mortality since this is an incident event. But apart from this difference in the nature of the health outcome event, cross-sectional studies are structured and analysed in the same way as cohort and case–control studies.

In environmental epidemiology, one of the principal uses of this study design is to assess the role of exposures that result in symptoms and biological changes, such as anaemia or decreases in lung function, but which do not necessarily cause individuals to seek medical care. In many instances, the biological changes are studied because they are early indicators of risk for mortality or the development of disease. Cross-sectional studies have proved effective for the study of chronic conditions such as hypertension, chronic obstructive lung disease and arthritis, as well as diseases that result in residual sequella such as stroke and coronary disease.

Cross-sectional studies are often undertaken in specific communities in response to concern about exposure from a point-source such as an industrial facility or hazardous waste site, or if specific health outcome rates appear to be increasing. For example, studies have been conducted in communities located close to industrial facilities. See Kreiss et al. (1981), Jaakkola et al. (1990) and Marttila et al. (1994), Dales et al. (1989), Ozonoff et al. (1987) and Baker et al. (1988), for examples relating to a DDT-manufacturing plant, paper mills, a natural gas refinery, and hazardous waste sites, respectively. This approach is useful if multiple possible exposures and a range of health outcomes must be examined. The findings of the cross-sectional study can then be used to design a more focussed cohort or case–control study.

Cross-sectional studies may also be based on national samples such as those obtained under the US National Health and Nutrition Examination Survey. For example, a study used data from a national health examination survey undertaken in the US to evaluate the association between lead exposure and hypertension by comparing the blood lead concentrations and blood pressure of the survey participants (Schwartz et al., 1986). Data from the survey were also used to examine the association between air pollution exposure and lung function (Schwartz, 1989) and to identify risk factors for childhood asthma (Gergen et al., 1988).

Design of cross-sectional studies

The key steps in conducting a cross-sectional study are:

- identify the source population;
- choose a sampling design and sampling frame for selecting the study participants;
- measure the exposure and health outcome status of the study participants.

Selection of study population

Since cross-sectional studies are generally used to study health outcomes for which identification of a population using mortality or medical records is not possible, the selection of a sampling frame is particularly important. Sampling frames include census lists of all households in a community, population lists based on holders of driver's licenses, voter registration files, medical care registrations, and other public records. If a sampling frame does not already exist, the investigators must create a sampling method that will yield a representative sample of the source population. In many community studies, investigators have found that the most complete coverage and highest participation rates are achieved when study staff go "door-to-door" to identify subjects and request their participation.

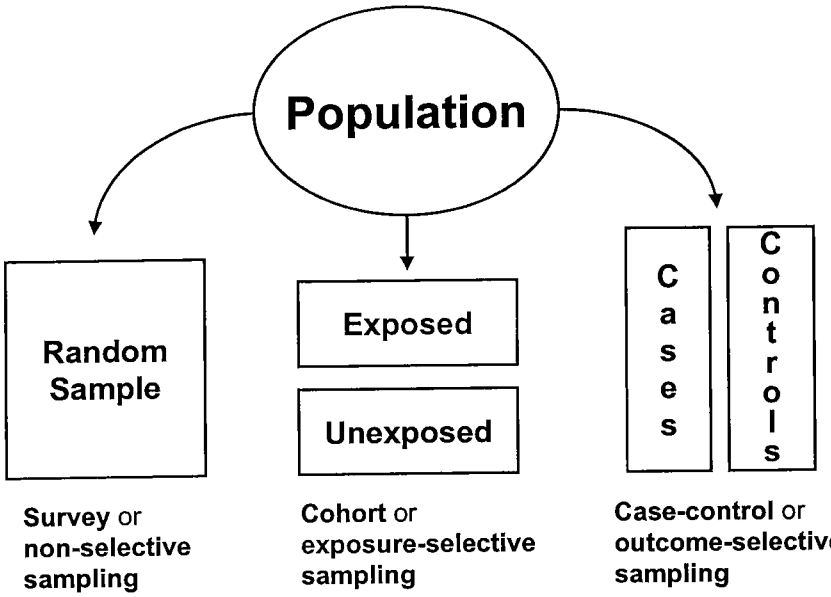
Study populations for cross-sectional studies can be selected using one of three strategies. Issues to be considered in deciding upon a strategy include the objectives of the study, the prevalences of the exposures and health outcomes, and the availability or non-availability of a sampling frame that could be used to identify either exposure or health outcome status (see Figure 6.7).

Perhaps the most common strategy is that of conducting a random sample of a population without regard to exposure or health outcome status, i.e. a **survey** or **census** sampling strategy. This approach is suitable for examining multiple possible exposures and health outcomes. The aforementioned cross-sectional studies which were based on the national examination survey data are examples of the survey approach. In these, the same study population was used to examine a number of exposure and health outcome associations. This strategy is also used when it is not possible to determine exposure or health outcome status before selecting the study population. For example, Sears and colleagues conducted a medical survey among 562 11-year-olds in New Zealand to evaluate the relationship between serum total IgE level, as an indicator of allergic diathesis, and airway hyperresponsiveness to inhaled methacholine (Sears et al., 1991). The investigators included in the study all children born at the regional maternity hospital eleven years earlier and who were still living in the province. This strategy was used because it was necessary to conduct the survey to determine both IgE levels and airway hyperresponsiveness.

Investigators may decide to invite all the members of a community to participate in a survey if the community is small or if it would be awkward to include some and to exclude others. The study by Kreiss and colleagues (1981), who invited all the residents of a small community located near a DDT manufacturing facility to participate in a medical survey, is an example of

a full community survey. The study found that the geometric mean level of total DDT in serum samples of the residents was 72.2 ng/mL compared to the national mean of 15.0 ng/mL. Within the community, total DDT-level increased with age and was associated with fish consumption and years of residence. The survey approach allowed the investigators to determine the average DDT levels in the community and to examine various risk factors associated with DDT serum concentration.

Figure 6.7. Sampling strategies for cross-sectional study



A second common strategy is for investigators to select the study population on the basis of **exposure status**. This strategy is used particularly when the study objective is to determine the risk of a specific environmental exposure. As with a cohort study, selecting subjects on the basis of exposure status is efficient if exposure prevalence in the source population is relatively small because the focus is on those individuals who contribute to the epidemiological analysis. In order to use this strategy, identification of subjects' exposure status, prior to their being selected for the study, must be possible. This is usually achieved in one of two ways — corresponding to the use of external and internal controls in cohort studies. Investigators may use existing exposure data to identify an exposed community and an unexposed community for comparison, or to identify various exposure sub-groups within a single community.

Jaakkola and colleagues used the external community comparison approach in their study of respiratory effects among residents of communities living near paper mills in Finland (Jaakkola et al., 1990; Marttila et al., 1994). They conducted a questionnaire survey in three communities that they had classified as severely, moderately, or nonexposed. They found that for the previous 12 months, residents in the severely exposed community were 11.7 times more likely to have reported eye symptoms and 2.0 times more likely to have reported nasal symptoms than residents of the nonexposed community.

The third strategy for selecting a study population is essentially a case–control design in that subjects are identified on the basis of their **health outcome status**. This is an efficient strategy if the proportion of the population with the health outcome is relatively small and can be used to analyse the association between prevalence of disease and environmental exposures. Subjects can be sampled on the basis of health outcome status if the relevant medical records are available. For example, medical records in a hospital clinic could be used to identify children with a diagnosis of asthma and a control group of children attending the same clinic who do not have asthma. These children could then be invited to participate in an examination to determine environmental risk factors for asthma. If existing records are not available, using a two-stage design to first screen the population to determine health outcome status, and then to conduct a more detailed study of cases and controls to determine exposure status, may still be efficient.

Assessment of health effects

The measure of health outcome occurrence in a cross-sectional study is prevalence, that is, the proportion of persons in the group who are affected by the health outcome at a point in time. As a practical matter, many cross-sectional studies examine the **period prevalence** rather than the **point prevalence** for outcomes such as symptoms. The point prevalence is the number of existing cases at a specified point in time, while the period prevalence is the number of cases during a specified interval of time. The cases include the existing cases at the beginning of the interval plus the incident cases during the interval. For example, a questionnaire may ask if the subject has experienced headaches or wheezing during the past three months, rather than simply asking the subject if he or she is currently experiencing such symptoms. Although there are subtle differences in interpretation, most investigators analyse and interpret these period prevalence measures as if they were measures of point prevalence.

Strengths and limitations

If a prevalent health outcome is being investigated, a cross-sectional study may be an efficient option since such a study usually costs less and requires less time to conduct than either an incident cohort or case–control study. Cross-sectional studies can be used effectively to study factors that do not change as a result of disease, such as ethnicity, blood group, genetic markers of susceptibility or external environmental exposures measured by objective methods. However, exposures measured by questionnaire or biomarkers may be affected by the simultaneous measurement of health outcome. Disease onset may influence subjects' recall of earlier exposures and may affect biological markers of exposure such as organochlorine pesticide concentration in adipose tissue. Thus cross-sectional measures of exposure can be used effectively only if verification that the measures are not affected by disease onset is possible.

Another major limitation of this type of study is that interpretation of cross-sectional associations between exposure and a prevalent health outcome may be subject to inaccuracy. For example, if persons with lung disease move to areas with low air pollution, a cross-sectional study would underestimate the association between air pollution exposure and lung disease risk. Similarly, if parents of children who had developed asthma reduced their cigarette smoking, a subsequent cross-sectional survey would be less likely to identify an association between the parents' earlier cigarette smoking and the onset of their children's asthma.

Given that the objective of most epidemiological studies is to determine the risk of an environmental exposure in causing a health outcome, the difficulty in interpreting the findings of a cross-sectional study with respect to risk of disease incidence can be considered a weakness.

Time-series and panel studies

Time-series and panel studies are not distinct types of study design, since they are based on cohort, cross-sectional and ecological studies. But their analysis and interpretation differ from those of the study designs already discussed. In panel and time-series studies, repeated observations of exposure and health outcome are made over time within the same study population. The analysis involves comparing variation in exposure status over time with changes in health outcome status over time. These study designs are most appropriate if the time interval between exposure and a variable health outcome is short (as in the case of respiratory symptoms or lung function), making it possible to determine whether a change in exposure is associated with a change in outcome.

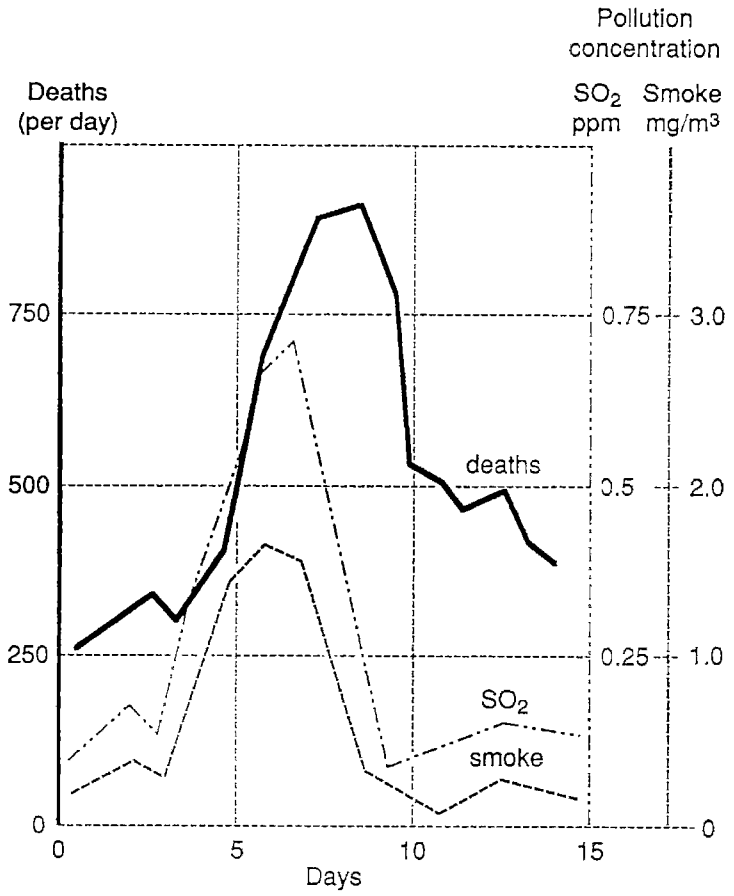
Most commonly, time-series studies are based on ecological or aggregate measures of exposure and health outcome, although similar analyses can be performed on individual observations of exposures or health outcomes. The term "panel study" is generally used to refer to studies involving repeated observations of individual subjects, although conceptually the design is the same as that of a time-series study in which the analysis examines the correlation between exposure status and outcomes over time.

Time-series studies

A time-series study based on aggregate data is essentially a temporal-comparison ecological study which examines an association between a variable exposure and a variable health outcome. Time-series studies have played an important role in air pollution epidemiology, for example in the investigation of the acute health effects of ambient pollutants. One of the best-known examples of a time-series study is that of the London smog episode of 1952 (United Kingdom Ministry of Health, 1954; WHO, 1991b) (see Figure 6.8.) This pollution event was particularly dramatic owing to the large number of excess deaths and the extraordinarily high levels of pollutants.

Subsequently, Schwartz and Marcus (1990) examined the association between daily mortality and "British Smoke" levels in London for the period 1958–1972. ("British Smoke" was measured in a specific manner and the exposure levels cannot be converted to modern units easily). Using autoregressive models, they found a highly significant relationship between mortality and particulate matter or sulfur dioxide, after controlling for temperature and humidity. Figure 6.9 demonstrates the exposure–response relationship between daily mortality and mean British Smoke level. Time-series ecological studies have also been used by investigators to study the effects of air pollution on mortality (e.g. Pope et al., 1992; Kunst et al., 1993); hospital admissions (e.g. Bates & Sizto, 1987; Bates et al., 1990; Pope, 1991) and emergency room visits for asthma (e.g. Bates et al., 1990). These studies have revealed an association between changes in ambient particulate concentrations and health outcome rates, even at concentrations which would generally not be considered harmful (Dockery & Pope, 1994). However, although time-series analysis of ecological data has been used very effectively in air pollution epidemiology, its application in other areas of environmental epidemiology has been limited. This is because large data sets with repeated measures of exposure are not generally available, and because the time interval between exposure and onset of health outcome is generally too long.

Figure 6.8. Mortality and air pollution in London, 1952



Source: Beaglehole et al., 1993; adapted from United Kingdom Ministry of Health, 1954.

Most time-series analyses are carried out using large, existing data sets on environmental variables, such as daily air pollution levels, and health outcome variables, such as daily hospital admissions. The investigators merge these data sets and then typically use regression techniques to control for lag periods (latency), autocorrelation, and the effects of factors (such as weather and temperature in air pollution studies), that co-vary with the exposure.

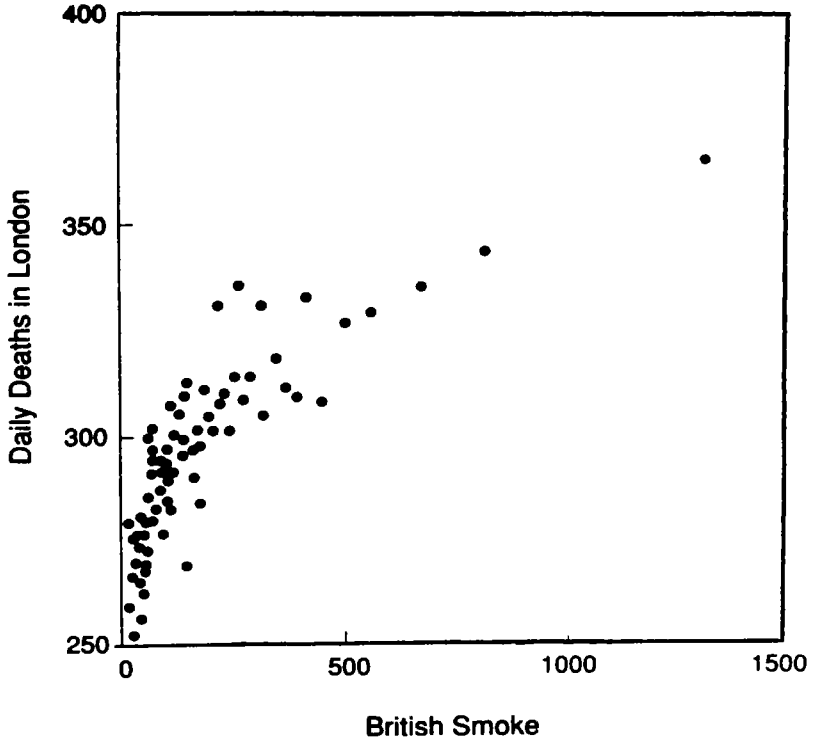
Panel studies

In a panel study, the investigator follows the subjects over time, measuring exposure and health outcome status repeatedly. Panel studies can be regarded as either cohort studies or time-series studies, depending on how the findings are analysed.

If exposure status is determined for each panel member and the panel is then followed over time to assess disease incidence or changes in health status, these studies are the same as cohort studies. The only difference is that repeated examinations are used to measure the outcomes. This approach is used to study health outcomes that are observed as a result of periodic medical examinations, rather than health outcomes such as death or hospitalization that can be studied

by using medical records. For example, a study of the incidence of newly-developed hypertension or diabetes could be undertaken by repeat examination of the subjects at regular intervals over time. Similarly, Sparrow et al. (1993) evaluated predictors of the new onset of wheezing among men by studying a cohort of 624 men included in the *Normative Aging Study*, who were examined at intervals of between three and five years over more than two decades. They found that current smoking, age, and postural heart rate change at the initial examination were predictive of the new onset of wheezing that occurred during the examination intervals.

Figure 6.9. Daily mortality and exposure to British Smoke in London



Source: Schwartz & Marcus, 1990.

The term "panel study" is more commonly used in environmental epidemiology, however, to refer to a study during which subjects are followed over time in order to measure the time-varying relationship between exposure and health status. These panel studies should be considered as individual-level time-series studies or even more appropriately as cohort studies that incorporate a time-series analysis. This approach has been used extensively to study the acute respiratory effects of air pollution (e.g. Kryzanowski et al., 1992; Romieu et al., 1992; Ostro et al., 1993). Generally, a panel of subjects is identified and asked to keep a daily diary over a period of several months to record symptoms, respiratory illnesses and medications used. The subjects are also requested to record their activities and daytime location (for example at work, school, home or outdoors). Some studies require in addition that subjects measure peak expiratory flow one or more times each day using a portable peak flow measuring device.

Measurements of daily exposure to air pollutants, such as ozone or sulfates, would also be collected from stationary air monitoring stations or personal monitors.

The analysis then examines the correlation between daily exposure levels and daily symptoms, illnesses, and peak flow levels. The analysis of these studies requires the use of techniques to control for the autocorrelation that occurs within subjects over time. For example, a person with an upper respiratory infection on one day is likely to have an upper respiratory infection on the following day; the two observations will not be independent.

Proportionate mortality/morbidity studies

In a proportionate mortality/morbidity study (PM study), the distribution of health outcomes in one study population is compared to the distribution of those same outcomes in another population. For example, the death certificates of all persons who died in a particular community near a hazardous waste facility might be examined to determine the proportion of deaths that was due to liver cancer. This proportion would then be compared to the proportion of deaths due to liver cancer that had occurred in the general population (the proportional mortality rate (PMR)). The distinguishing feature of this type of study is that no information is obtained directly with respect to the characteristics of the whole source population. This is because all of the subjects have experienced a health outcome. This study type is therefore fundamentally different from other epidemiological study designs. Evaluation of the relative proportion of the health outcomes is only possible through comparison of the two series of outcome data. Because a PM study compares subjects who have experienced an outcome of interest (i.e., a specific cause of death or morbidity) and other health outcomes, some investigators have suggested that PM studies should be considered variants of case-control studies (Rothman & Greenland, 1998). The analytical methods (i.e., use of the odds ratio, as discussed in Chapter 7) are similar.

A PM study is typically carried out as the initial step in an outbreak investigation, if the focus of concern is apparent excess mortality or morbidity. This type of study may be inexpensive and quick to complete if data on mortality or morbidity are readily available. For example, death certificates of all persons who died in a particular region may be obtainable from a governmental agency. The availability of such data means that neither enumeration nor sampling of the actual source population is necessary.

Nevertheless, this study design has substantial weaknesses which can limit a study's validity and interpretation. For instance, death certificates or morbidity data may be incomplete and thus fail to represent the health outcomes of the source population accurately. This might be the case if some members of the source population, for example, the unemployed or poor, had less access to medical care than other members, with the result that their deaths were "missed" or their causes of death misclassified. Bias can also occur if the medical practitioners in the community are more or less likely than medical practitioners in the surrounding region to ascribe cause of death to a particular cause, such as malignant mesothelioma. Problems can also arise if the various study populations are not comparable with regard to age, gender or other characteristics. Unfortunately, information on potential confounders is rarely obtainable from the records upon which PM studies are based. Moreover, even if the populations are comparable, and causes of death are complete and correctly classified, a PM study may still encounter difficulties. This is because interpretation of the PMR as an estimate of the relative risk assumes that the two

populations have the same total mortality rate. But clearly this is unlikely in situations in which an excess risk of mortality due to some specific causes of death has been observed; an increased rate for some causes of death would increase the total mortality rate.

The approach of a PM study is to obtain information on a series of health outcomes, such as deaths from death certificates, for a population of interest, such as persons who died while living in a highly polluted region. Similar information would be obtained on the health outcomes occurring in another population, e.g. deaths among residents living in an unpolluted region. The analysis would then compare the distribution of causes of death as recorded on the two series of death certificates: for example, to evaluate whether proportionately more deaths due to lung cancer were recorded on death certificates obtained from the polluted region, than on those obtained from the unpolluted region. More specifically, investigators could obtain death certificates for all persons who died during a five-year period in a town located near an industrial facility. They would then compare the causes of death among town residents with the distribution of causes of death in a surrounding unpolluted region. If 18.9% of deaths among town residents were due to lung cancer, compared to 6.3% of deaths among persons in the surrounding unpolluted region, the PMR would equal 3.0. A similar type of analysis could be carried out for cancer incidence if the town and surrounding region were included in a population-based tumour registry and cancer incidence data available.

Experimental studies

Experimental studies differ fundamentally from observational studies since the investigators determine who will be exposed. A well-designed human experiment is perhaps the most convincing means of demonstrating a causal relationship between exposure or prevention of exposure, and a health outcome. Experimental studies are not common in environmental epidemiology though, since most exposures that are of interest are potentially detrimental, rather than beneficial. A **controlled human exposure** study can be undertaken to measure the effects of a potentially toxic substance on human subjects. But such studies are generally considered to be clinical rather than epidemiological studies since they typically take place in medical research facilities, and include relatively few subjects.

Randomized controlled trials are experimental studies that are undertaken to evaluate a new preventive or therapeutic regimen (Beaglehole, et al.,1993; Lilienfeld & Stolley, 1994). Subjects who already have a disease or who are at high risk of developing a disease are randomly allocated to groups which are then assigned different treatments. Another group — the comparison group — receives no treatment or an already standard treatment. After treatment, the groups are followed to determine whether the experimental treatment affects the health outcome over time. **Field trials** are similar to randomized controlled trials, but they involve high-risk, disease-free subjects who are selected from a community and studied "in the field" rather than in a hospital.

Randomized trials and field trials have been used effectively to assess the value of new therapies in treating acute diseases of environmental origin. For example, a study was undertaken during a cholera epidemic in Bangladesh in 1983 to compare rice-based and glucose-based oral rehydration solutions for treating diarrhoea (Molla et al., 1985). Similarly, vaccination trials have been used to assess the efficacy of new vaccinations in preventing communicable diseases.

Environmental interventions are often applied to groups of subjects or communities. A **community trial**, in which communities rather than individuals are randomized for the intervention, may therefore be more appropriate. Community trials have included studies of communities to compare the effectiveness of different types of sewage treatment and water chlorination in preventing diarrhoeal disease, and of the addition of different fluorine concentrations to drinking-water supplies to prevent dental caries. A study of children living in New York in communities that were served by water supplies to which sodium fluoride had been added is a classic example of this approach (Asp & Schlesinger, 1956). The study observed a substantial decrease in the incidence of decayed, missing and filled teeth among children in the treated community during the ten years following the onset of water treatment (see Table 6.3).

Table 6.3. Permanent teeth decayed, missing, or filled (DMF) per 100 children, ages 6–16, based on clinical and roentgenographic examination ten years after start of fluoridation — New York, 1954–55

Age	Number of children		DMF teeth per 100 children		
	Newburgh (treated)*	Kingston (comparison)	Newburgh (treated)*	Kingston (comparison)	% difference (N-K)/K*100
6–9	708	913	98.4	233.7	-57.9
10–12	521	640	328.1	698.6	-53
13–14	263	441	610.1	1170.3	-47.9
15–16	109	119	975.2	1648.7	-40.9

* Sodium fluoride was added to Newburgh's water supply beginning May 1945

Source: adapted from Asp & Schlesinger (1956) and Lilienfeld & Stolley (1994).

Design of experimental studies

The essence of an experimental strategy consists in selecting two or more groups of subjects randomly so that they are as comparable as possible (Puska, 1991; Lilienfeld & Stolley, 1994). The exposure to the factor under study is then assigned to one of the groups (usually a treatment group), but not to the other (the comparison group), and the health outcomes for the groups then compared. The necessary steps are:

- identify a sampling frame, such as a roster of patients in a clinic;
- select potential subjects based on explicit eligibility criteria;
- request the subjects' participation in the study;
- randomly allocate the consenting, eligible subjects into groups.

Random allocation or randomization of subjects is a key feature of a well-designed experimental study. The study groups must be as comparable as possible with respect to any factor that could influence the health outcomes. Random allocation is undertaken to minimize

the potential for confounding or selection bias. Many techniques have been recommended for performing random allocation. Random number tables, for example, can be used to ensure that treatment assignment is random and not influenced by the subjects or investigators. But whatever method is used, the investigator should measure the group's key characteristics in order to verify the comparability of the groups.

Once the study groups have been allocated, the experimental intervention is administered to one or more of them. If observation bias is to be minimized, neither subjects nor observers should know who receives the intervention. Subjects and observers should therefore be "blinded" with respect to the subjects' treatment status. In studies of clinical treatment or preventive regimens, subjects in the comparison group are commonly given a placebo which appears to be identical to the active treatment. Similarly, in experimental studies of environmental exposures, subjects are not informed as to when the experimental exposure will occur. Thus in experimental studies of indoor air pollution, investigators have manipulated the ventilation rate and amount of fresh air supplied to office buildings and requested the subjects to record their symptoms and health complaints in daily diaries (see, for example, Menzies et al., 1993; Jaakkola et al., 1994). The subjects did not know when the ventilation rates would be changed, which meant that this information could not influence their tendency to report symptoms.

Table 6.4 presents some findings of a study by Jaakkola and colleagues of the experimental manipulation of air supply (70% recirculated air versus 0% or all fresh air) in two identical buildings (Jaakkola et al., 1994). The proportion of return air was altered in the buildings in a blinded, four-period crossover trial with each period lasting one week. The building occupants were not informed of any changes made to the building's ventilation and were requested to report symptoms in a daily diary. The investigators found that the symptom scores of the 75 subjects for the index (recirculated air) phases and reference (all fresh air) phases did not differ substantially.

An intervention is typically administered for a fixed period of time and the subjects or communities then followed to determine health outcome status. The methods for following up the study groups of experimental studies are similar to those used for following up the study groups of cohort studies.

A **cross-over design**, in which each subject serves both in the exposed group (at times of intervention) and the unexposed group (at other times), can also be applied in experimental studies. The health outcome status of each subject is compared during the exposed and unexposed phases. This was the approach taken in the indoor air studies cited above (Menzies et al., 1993; Jaakkola et al., 1994). Its advantage is that each subject serves as his or her own comparison, thereby minimizing the potential for confounding due to non-comparability. However, this approach can only be used effectively for studying rapidly reversible health effects with clear temporal relationships to the putative exposure. As with the indoor air studies, the cross-over design can be useful for studying symptoms and subjective conditions, provided subjects record symptoms while they are blinded to their exposure status.

Table 6.4. Probability ratio of having more or fewer symptoms with 70% recirculated air compared to all fresh air ventilation in office buildings

	Occurrence during 70% air recirculation				
				Probability ratio	
Outcome	No. with more symptoms	No. with fewer symptoms	No. with equal symptoms	Point estimate	95% CI
Mucosal irritation score	25	30	17	0.83	0.49, 1.42
Eye symptoms	13	19	40	0.68	0.33, 1.39
Nasal dryness	22	15	35	1.47	0.76, 2.83
Nasal congestion	20	18	34	1.11	0.59, 2.10
Pharyngeal irritation	18	20	34	0.90	0.47, 1.72
Skin reaction score	21	21	30	1.00	0.58, 1.91
Skin symptoms	20	18	34	1.11	0.59, 2.10
Rash	10	9	53	1.11	0.46, 2.70
Allergic reaction score	23	33	16	0.70	0.41, 1.18
Nasal excretion	14	19	40	0.74	0.38, 1.46
Sneezing	12	26	34	0.46	0.24, 0.90
Cough	7	20	45	0.35	0.15, 0.79
General symptom score	26	22	24	1.18	0.67, 2.08
Headache	25	16	31	1.56	0.84, 2.89
Lethargy	11	12	49	0.92	0.42, 2.00
Perception of air quality					
Unpleasant odour	9	19	44	0.47	0.21, 1.05
Stuffiness	19	19	44	1.00	0.53, 1.89
Dustiness	13	16	43	0.81	0.34, 1.91

Source: Jaakkola et al., 1994.

Strengths and limitations

Experimental studies are the most definitive study design since the investigator assigns the exposure. Experimental studies also make it possible to evaluate the effect of dose and study temporal relationships since exposure or treatment is assigned. However, experimental studies are limited in that they can be used only for the study of short-term, reversible effects of potentially harmful exposures, or of treatment, or of preventive interventions. Also, because experimental studies require follow-up, they share some of the limitations of cohort studies, namely, potentially high cost and loss to follow-up.

The community trial is an effective experimental study design if the problem to be studied could be affected by community-level intervention. Community trials do raise some concerns, however. In common with ecological studies, it is not possible to determine whether group-level associations reflect individual-level effects accurately. Ethical issues of informed consent must also be considered before undertaking a community trial; requesting permission to administer the treatment from each community member is not always feasible. Finally, as individuals living in the same community tend to be relatively homogeneous, care should be taken to ensure that the statistical analysis of a community trial is adjusted for correlation within groups.

Community trials are also limited in that often it is not possible to assign treatment randomly. For example, it may not be possible for an investigator to randomly assign which community treats its water supply with sodium fluoride. If subjects or communities are not assigned randomly, the potential for confounding in the study increases.

6.3 Methodological issues in study design

As discussed in Chapter 3, the goal of any environmental epidemiology study is to obtain an accurate estimate of the quantitative association between an environmental factor and disease occurrence. The basic strategy for enhancing the quality of a study is to design the study to minimize the potential for error, while being aware of considerations such as cost and efficiency. Potential sources of **systematic error** (bias) and **random error** are reviewed in Chapter 3. There are three general forms of bias that have been distinguished for epidemiological studies - confounding, selection bias, and information bias (Rothman & Greenland, 1998). Information bias can be minimized by paying attention to quality assurance principles in the collection and analysis of the data. Confounding and selection bias can be controlled while designing the study and may be adjusted for during the data analysis - as discussed in Chapter 7.

Control of confounding at study design stage

Control of confounding at the design stage can be carried out by randomization, restriction or matching. **Randomization** 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. This approach has some conceptual and computational advantages, but can limit the number of potential study subjects and the informativeness of the study.

But even without full restriction, eligibility criteria for subjects are generally selected so that the study population is to some extent restricted, thereby facilitating analytical control of confounding and effect modification. A case–control study of lung cancer and radon exposure may be restricted to subjects more than 45 years of age since the number of cases below that age may be too few to be interpretable. Likewise, in a cross-sectional study of organochlorine pesticide concentrations in breast milk, investigators restricted the study population to women between 20 and 35 years of age, who had just given birth, with parity less than three (Baker et al., 1991). Age and parity, as an indicator of prior breast feeding, were considered to be potentially confounding variables. The decision to restrict the study population by age and parity

was made in order to exclude the small proportion of women of high or low age and high parity. The number of such women would have been too small to interpret or analyse effectively.

A third method of control is to **match** study subjects on potential confounders (e.g. matching for age, gender and ethnicity). This will eliminate confounding from a cohort study, but can be expensive and is therefore rarely carried out. Furthermore, information on potentially confounding variables is usually not readily available when identifying cohorts.

Matching can also be carried out for case–control studies, but it will not remove confounding — merely facilitate its control in the analysis — and is expensive. Furthermore, matching may reduce precision in a case–control study if it is performed on a factor which is associated with exposure but which is not a risk factor for disease (and hence not a true confounder). However, matching on a strong risk factor usually increases precision, and may also have practical advantages.

In case–control studies, matching 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 performed on a case by case basis, which is known as **individual matching**; each control subject is uniquely linked to a case, although there may be more than one control per case. Matching can also be performed for groups of subjects. This is **frequency matching**. The distribution of the relevant risk factors, such as proportion of subjects who are male and proportion who smoke cigarettes, is measured for the cases. Controls subject to the same frequencies of these factors are then selected.

Population sampling

While most textbooks emphasize the difference in sampling units between individual and ecological studies, the distinction is often not so clear in environmental epidemiology studies. Exposure often occurs on a community-wide basis (e.g. air pollution). Individuals of the same community may therefore be assigned identical exposure values. This is an ecological study approach to establishing individual exposures. In the same study, health outcomes are determined for individuals. Such studies are generally analysed as individual-level studies, but they share many of the characteristics, and limitations, of ecological studies. For example, exposures are almost never truly homogeneous within a community. Furthermore, since so many "individuals" are assigned exactly the same exposure, the statistical analysis and interpretation can be affected as the true distribution of exposures is not known.

Problems can also arise when, as is often the case in environmental epidemiology studies, the investigator selects participants for community studies on the basis of households, rather than individuals. Thus it is the household which is the sampling unit and not the individual. Since household members of the same household may be more similar to one another than to other individuals, the analysis should account for any resultant covariation. This complication can be avoided altogether by including only one member from each household in a community study.

Methods for sampling persons from populations have been described in several textbooks (see, for example, Kelsey et al., 1996; Kahn & Sempos, 1989) and are reviewed in Chapter 3, pp.61. While a simple random sample may appear to be the best means of obtaining a representative

sample, it is usually more efficient to obtain stratified, systematic, cluster, or multi-stage samples. Stratified sampling is more efficient than simple sampling when the distribution of subjects according to relevant variables, such as age or gender, is not uniform. Under most circumstances, study precision can be increased by using stratification whenever the population can be subdivided into sub-groups that are more homogenous with respect to the variables of interest, than is the population as a whole (Kelsey et al., 1996).

One of the advantages of systematic sampling (for example, sampling every second person) is that the investigator does not need to determine the entire sampling frame of potential subjects before beginning the sampling. Moreover, under field conditions systematic sampling is often simpler to implement than random sampling. For example, sampling births in a hospital by making a systematic sample of every fifth birth would be easier than applying a random selection procedure for each separate birth, with a one-fifth possibility of that birth being selected for recruitment. The disadvantage of systematic sampling though is that identifying patterns in the sampling frame, which could render the systematic sample unrepresentative, can be difficult. Selecting sampling intervals that are not related to exposure or health outcome is clearly important. Thus sampling of births in a hospital should not be based on day of 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 can be difficult unless an assumption can be made that a systematic sample from a randomly ordered population is the same as a simple random sample. Most studies make this assumption and analyse systematic samples as if they were simple random samples.

Sample size and precision

Deciding how many subjects to include in the sample is central to any epidemiological study. The sample of subjects should be large enough to ensure that the hypothesized effects can be estimated and with adequate precision, given the anticipated variation in the key variables to be analysed. As noted in Chapter 7, statistical analysis aims to estimate the precision of observed effect 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. 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 that must be included to achieve the desired statistical precision.

Several textbooks have presented formulae and tables for calculating the number of subjects required to achieve the desired precision for various types of epidemiological study and statistical analysis (see, for example, Breslow & Day, 1980; Breslow & Day, 1987; Kahn & Sempos, 1989; Lwanga & Lemeshow, 1991; Selvin, 1991).

6.4 Strategy in choosing a study design

In order to choose an appropriate study design, the investigator must appreciate the relative strengths and limitations of the various study designs, which are discussed above in Section 6.2 for each of the major types of study designs. After evaluating the strengths and limitations of the study design options, the investigator must consider several other factors before choosing a study design. These include the purpose of the study, the extent of exposure, the nature and frequency of the disease occurrence, and the availability of existing records of exposure and health outcome occurrence.

Descriptive studies are chosen when the purpose is to explore unknown but potential cause–effect relationships for the purpose of generating hypotheses. They are also used to investigate in order to quantify environmental health impacts when investigating known cause–effect relationships.

Analytic studies are often motivated by concern about a specific environmental factor or disease occurrence. The investigator can use a cohort design if the motivation for the study is "exposure-oriented" and a case–control design if it is "disease-oriented", although the association between an exposure and disease occurrence can be evaluated using either study design. If the association between a specific environmental exposure and the occurrence of multiple diseases is being examined, a cohort design is appropriate, but if the association between a specific disease and multiple environmental exposures is being examined, a case–control design would be preferable.

If an analytical study design is favoured, an exploratory analysis using existing records to develop more focussed hypotheses can be undertaken. Indeed, a series of studies is often needed to refine and evaluate hypotheses about the role of environmental factors in disease occurrence. Investigators generally begin with ecological or cross-sectional studies before considering cohort or case–control studies. An example is the series of studies undertaken to determine the cause of asthma in Barcelona (Antó et al., 1989; Antó & Sunyer, 1990). Approaches used to evaluate other environmental poisoning episodes have been described in material prepared by the World Health Organization (see, for example, WHO, 1991b; WHO, 1992c).

When evaluating the incidence of a health outcome, use of a case–control design to sample from the study base is usually more efficient than use of a cohort design to examine the entire study base. Admittedly, sources of bias can prove harder to identify if a case–control strategy is used, but for many environmental studies, the distinction between cohort and case–control designs regarding potential bias is not pertinent. Environmental exposures commonly occur among dynamic populations. Therefore, a case–control study which samples from the experience of a dynamic population may be as feasible as a cohort design, for which identifying the exact source population would also be difficult because thee is no readily available sampling frame.

The incidence rate of virtually all diseases in any population tends to be rare so that a case–control design is generally an efficient strategy. However, such a strategy may be ineffective if the exposure prevalence is also rare. There are four primary strategies for assessing the effects of rare exposures on rare health outcomes, none of which is entirely satisfactory.

One option is to conduct the study with a very large source population. However, this is feasible only if existing databases can be used. Unfortunately, few data sets exist that include adequate

information on both exposure and disease occurrence among appropriate populations. Most commonly, use of existing data requires an ecological analysis.

An alternative would be to undertake a case–control study using a very large study population. But this would only be worthwhile if the gathering of information on exposure status from the study subjects (for example, via. mail-back or phone-administered questionnaire) was inexpensive. Moreover, this approach is used less frequently than formerly, owing to increased awareness of the importance of obtaining valid measures of exposure, which generally require direct environmental or biological sampling.

A third strategy would involve restricting the study population to a "high-risk" population, that is, a population with a high proportion of exposed persons or high levels of exposure, or a population of individuals of greater susceptibility. This strategy is applied in studies of occupationally exposed populations which are undertaken in order to assess the potential effects of a substance found at lower concentrations in the general environment. However, this approach may limit the size of the population available for study and therefore restrict the generalisability of the findings.

A cohort study design can be an efficient means of evaluating the effect of a rare exposure. However, this design is inefficient in ascertaining disease occurrence. If an historical cohort study is conducted, time and cost may be decreased, but records are crucial for identifying the study population and measuring past exposure. Thus historical cohort studies rarely prove to be feasible for environmental health studies.

Chapter 7

Analysis and Interpretation of a Study

Learning objectives

- Understand the basic measures of effect used in environmental epidemiology studies.
- Understand how to apply the basic statistical methods for data analysis.
- Be able to propose simple methods for controlling confounding during the analysis of the data.
- Be able to recognize when more advanced methods of statistical analysis are required.

7.1 Approach to data analysis

This chapter provides an overview of the analysis of environmental epidemiology studies. Data analysis, in common with other aspects of environmental epidemiology such as exposure assessment and health outcome assessment, may be conducted primarily by professionals with specialized skills (for example, biostatisticians), but will nevertheless require input from all members of the study team. Thus an epidemiologist should be involved in the data analysis, just as a biostatistician should be involved in the study design.

As discussed in Chapter 6, the usual approach in an epidemiological investigation is to select a study population and measure disease occurrence in the population, particularly in relation to exposure status and other possibly relevant factors such as age or gender. Data analysis makes it possible to express quantitatively the disease occurrence measures and effect measures (defined in Chapter 3) relating exposure and health outcome variables. Data analysis also enables estimation of the precision of these measures, given the possibility of random error. Stratification and modeling are used to identify and adjust for confounding factors that could result in systematic error (bias) if not taken into account.

Effect estimation

Data analysis in analytic and experimental studies aims at estimating the effect of a particular exposure (or demographic factor) by comparing the occurrence of disease in those exposed, with the occurrence of disease in those not exposed. **Effect estimation** usually involves combining the data with a statistical model. In so doing it can be ascertained whether the disease risk differs between the exposed and non-exposed groups. The size of the effect of exposure and the precision of this effect estimate can also be calculated. The definitions and formulae for calculating measures of effect were presented in Chapter 3.

In many instances, full understanding of the effect of an exposure on a disease outcome may require more than calculation of a crude measure of effect such as the rate ratio. Determining

whether the effect of exposure differs among the different sub-groups of the study population may also be useful. For example, Katsouyanni et al. (1993) studied the interaction between air pollution and high temperature in relation to the excess mortality that had occurred during a major heat wave in Greece in July 1987. They found that the increase in number of deaths in Athens (where air pollution levels are relatively high) was much greater (97%) than in other urban areas (33%) and in non-urban areas (27%). Further analyses suggested that the threshold of effect for various air pollutants appeared to be lower on extremely hot days than on less hot days. Such **effect modification** occurs when the estimate of the effect of exposure is associated with the level of another factor in the source population (see pp.52).

Since **dose–response** is an important criterion to consider when interpreting whether an observed association may be causal, it is generally important to evaluate whether effect estimates change systematically by a gradient of exposure magnitude or dose. The most common analytical approach is to compare sub-groups of the study population, characterized by different levels of exposure, to a non-exposed group.

When studying the long-term health effects of environmental exposures the **time patterns** of the study exposure must also be considered (Pearce et al., 1988). This does not affect use of the statistical methods presented here; rather it affects the definitions of exposure that are used in the statistical calculations. In particular, if a person accumulates exposure over a period of time, this must be allowed for in the analysis. Thus each person in the study "moves through" cumulative exposure categories as he or she accumulates exposure over time, just as each person "moves through" different categories of confounding factors such as age (Pearce, 1992).

Confounding and bias

An essential purpose of data analysis is to identify and control for extraneous factors that may distort the association between an exposure and the health outcome. **Confounding** can be controlled by stratifying the data into sub-groups according to the levels of the confounder(s) (for example, heavy smokers, light smokers, non-smokers), and calculating a summary effect estimate by aggregating the information obtained from the data across the strata. It is usually not possible to control for more than two or three factors simultaneously in a stratified analysis since otherwise the data may become sparse in one or more strata. Multivariate modeling methods allow the simultaneous control of a greater number of confounders and may therefore be preferable. 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.

Stratification can also be used to control partially for the effects of selection and information bias if surrogate variables can be identified that are correlated with the sources of bias. Even if such analytical control is not possible, using strategies such as analysis of repeated measurements to estimate quantitatively the *validity* and *reliability* of the study observations (see Section 3.3), is usually worthwhile.

7.2 Basic statistics

This section briefly discusses the fundamentals of statistical analysis. It builds on material presented in the WHO *Basic Epidemiology* text (Beaglehole et al., 1993). The focus here is on methods for categorical data, since this is the type of data most commonly encountered in environmental epidemiology.

Statistical distributions and summary measures

Measurement scales are used to quantify the variables to be considered in a statistical analysis. **Nominal scales** classify observations into categories, such as gender or ethnicity. **Ordinal scales** assign rank orders to categories, such as mild, moderate or severe symptoms. **Interval scales** measure continuous variables in which the distance between two measurements are defined, but for which there is no absolute zero point or origin (e.g. intelligence quotient). Conversely, in **ratio scales**, the distance between two measurements is defined and there is a logical zero point (e.g. carbon monoxide concentration in air). The existence of a zero point means that the ratio between the two measures can be defined. For example, one metre is twice as long as 50 cm. Similarly, it may be possible in an epidemiological study to observe, for example, that the incidence rate of a disease among an exposed population is twice as high as that in an unexposed population. This ratio was defined in Chapter 3 as the rate ratio, or more generally the relative risk.

Data distributions can be summarized in various forms, including frequency tables, histograms, bar charts, cross-tabulations and pie charts. However, a summary **measure of central tendency** is usually helpful too. The **mean** (or average) is the most commonly used measure of central tendency for continuous variables because of its convenient statistical properties and its straightforward interpretation when the variable is symmetrically distributed. The mean in a population can be estimated as the sum of the observations divided by the number of observations (n):

$$\bar{x} = \frac{\sum x_i}{n}$$

Once the data have been summarized, a particular data distribution may be assumed in order to obtain a summary **measure of variability** of the data. The most useful data distribution assumed for continuous variables is the **normal distribution** (also called Gaussian distribution). This is symmetrical around its mean and unlimited both downwards and upwards (commonly described as a bell-shaped curve). The most common measure of variability is the **standard deviation**. The standard deviation is especially useful when the underlying data distribution is approximately normal. If data are not normally distributed, they can often be made approximately normally distributed by an appropriate transformation (e.g. a logarithmic transformation for data distributions that are log-normal):

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Generally it is not possible to study the entire population in which one is interested. A random sample must be taken instead and its characteristics related to the total population. If repeated samples are taken from the same population, the mean will vary between samples. Even if the underlying data are not normally distributed, the means of the samples will be approximately normally distributed provided that the samples are sufficiently large. (How "large" will depend on the extent to which the data are non-normally distributed). The standard deviation of the sample means is termed the **standard error** of the mean. If the means are approximately normally distributed, about 95% of sample means will lie within 1.96 standard errors of the overall population mean (see Beaglehole et al., 1993). Most studies involve only one sample, but the standard error can be estimated by dividing the standard deviation of the sample by the square root of the number of subjects contained in the sample:

$$SE = \frac{SD}{\sqrt{n}} \quad \text{(for normal distribution)}$$

Most environmental epidemiological studies involve categorical rather than continuous data; specifically, health outcomes are dichotomized into whether or not disease occurred. This necessitates the calculation of a proportion (p). Although the occurrence of disease is categorical (i.e., it did or did not occur in an individual), the proportion is a continuous variable on a ratio scale. Another feature of a proportion is that its values must be between 0 (none affected) and 1 (all affected), inclusive. For example, the proportion of births subject to congenital malformation might be estimated for a particular geographical area, for a particular period of time. Another example would be the proportion of exposed persons dying during a specified time interval. Under the **binomial distribution**, the sampling distribution will approximate to the normal distribution with mean (p) and standard deviation $\sqrt{p(1-p)}$ if the sample is sufficiently large. Thus, the proportion who die (i.e. the mean score for a population in which a death scores 1 and a survivor scores 0), and the standard deviation of this proportion (i.e. the standard error of the mean score) can be calculated:

$$s = \sqrt{\frac{p(1-p)}{n}} = \frac{[b(n-b)]^{0.5}}{n^{1.5}} \quad \text{(for binomial distribution)}$$

Poisson distribution is another statistical model that is particularly important in analysis of epidemiological data (Ahlbom, 1993). The Poisson distribution reflects the probability of a certain number of events occurring in one time interval when the number of time intervals is extremely large but the chance of an event occurring in any one time interval is small (Kleinbaum et al., 1988; Ahlbom, 1993). The formula is applied using the person-time experience of the study population rather than persons as the unit of observation. The Poisson

distribution is a discrete distribution since only an integer number of events can occur in any one time interval. The distribution is limited downward by 0 but unlimited upward. A useful property of this distribution is that the measure of central tendency (mean rate) and measure of variability are directly related; specifically, the mean and variance (square of the standard deviation) are equal. Thus under a Poisson distribution, if "b" events occur during a "Y" time interval, the mean rate would be equal to b/Y, the standard deviation would be the square root of b/Y and the standard deviation of the mean rate (standard error) can be calculated :

$$s = \frac{b^{0.5}}{Y} \qquad \text{(for Poisson distribution)}$$

These three statistical models for data distribution form the basis of most of the statistical analyses presented in this chapter.

Testing and estimation

Once a sample has been selected, it can be used to estimate the characteristics of the underlying population. The most common approach is to construct a **confidence interval** around the estimate, i.e. to construct a range of values surrounding the estimated mean which has a specified probability of including the true population mean. The usual practice is to use 90% or 95% confidence intervals, but these values are arbitrary (see Beaglehole et al., 1993). For normally distributed data, the 95% confidence interval for the true population mean is:

$$\bar{x} \pm 1.96 \times SE(x)$$

where $\bar{x}$ is the mean of the sample and SE(x) the standard error, estimated from the standard deviation of the sample divided by the square root of the sample size. This confidence interval depends on two quantities ($\bar{x}$ and SE(x)) which are estimated from the sample itself. Different results will be obtained from different samples. However, for 95% of the time, provided that the samples are sufficiently large, the confidence interval estimated from the sample will contain the true population mean. The variability of the individual data can be estimated by the **tolerance interval**, which is calculated with the same formula as the one above, replacing the standard error with the standard deviation.

Many environmental epidemiology studies aim to measure the difference in disease occurrence between groups exposed and not exposed to a particular factor. For example, if the proportion of pregnancies subject to congenital malformation was estimated for a geographical area where nitrate levels in drinking-water are high, a comparison would then be made with the corresponding proportion for an area where nitrate levels in drinking-water are low (or with the corresponding proportion for all births nationally). A test statistic (z) can be used to calculate the **p-value**, which is the probability that a difference as large as the one observed (e.g. in the proportion of congenital malformations observed for the two different populations) had occurred by chance, and assuming the **null hypothesis**, that is that there was in fact no actual difference between nitrate levels and congenital malformations and no uncontrolled confounding or bias:

$$z = d/SE(d)$$

Where d is the observed difference and SE(d) the standard error of the difference, the p-value can be calculated for any value of z by using standard statistical tables (Rosner, 1986).

In the past, p-values were often used to describe the results of a study as "significant" or "not significant", on the basis of an arbitrary level of significance (e.g. p = 0.05, probability of chance result = 5%). However, it is now recognized that there are problems with this approach (Rothman & Greenland, 1998). First, the p-value associated with a difference in outcome between two groups depends on two factors: the size of the difference and the size of the study. A very small difference may be statistically significant if the study is very large, whereas a very large difference may not be significant if the study is very small. P-values thus combine two phenomena which should be kept separate: the size of the effect estimate and the precision of the effect estimate.

A second problem with significance testing is more fundamental. The purpose of significance testing is to reach a decision. However, in environmental epidemiology research, decisions should be based not on the results of a single study, but on information from all available studies, as well as on non-statistical considerations such as the plausibility and coherence of the effect in the light of current theoretical and empirical knowledge.

The problems of significance testing can be avoided, however, if it is understood that the principal aim of an individual study should be to quantitatively measure the occurrence of disease or the effect of an exposure (the **point estimate**), rather than just to decide whether an effect is present. The point estimate should be accompanied by a confidence interval which indicates the precision of the point estimate by providing a range of values within which it is most plausible that the true effect may lie (Gardner & Altman, 1986). Thus the point estimate reflects the size of the effect whereas the confidence interval reflects the study size on which this effect estimate is based.

7.3 Basic data analysis

The health outcome being studied is sometimes a continuous measure (e.g. blood pressure, change in lung function). Most continuous outcome variables are normally distributed or can be made approximately normally distributed by a suitable transformation. Standard statistical approaches for normally distributed data can therefore be used. In the majority of environmental epidemiology studies, however, the health outcome of interest is dichotomous; i.e. the study participants are classified as to whether or not they experience a particular health outcome (e.g. lung cancer, respiratory disease). In this text, emphasis is given to methods that are applicable to dichotomous outcomes, resulting in the disease occurrence rates and measures of effect defined in Chapter 3. Table 7.1 summarizes the various study types and corresponding analytical methods.

Initial exploration of data and descriptive analysis

The initial descriptive analysis uses frequency distributions, summary measures such as the mean of a continuous variable or disease incidence rate, and graphical methods such as a histogram to examine the characteristics of individual variables. Quantitative methods can be used to test whether a data distribution is closer to normal than log-normal or some other model (e.g. the Kolmogorov-Smirnoff test). The data distributions of various variables are then compared to identify associations that necessitate stratified or multivariate analysis.

Table 7.1. Study designs and corresponding analytical methods for epidemiological studies involving dichotomous outcomes

	Cohort study (rates)	Cohort Study # (risks)	Case–control study *
Numerator	persons (cases)	persons (cases)	persons (cases)
Denominator	person-years	persons	persons (controls)
Occurrence measure	rate (incidence rate)	risk (incidence proportion or cumulative incidence)	odds
Effect measure	rate ratio (incidence density ratio)	risk ratio (cumulative incidence ratio)	odds ratio
Stratified analysis	Mantel-Haenszel (also SMR, SRR)	Mantel-Haenszel (also SRR)	Mantel-Haenszel
Multivariate analysis	Poisson regression @	binomial regression	logistic regression

Notes: SMR = standardized mortality ratio, SRR= standardized rate ratio, standardized risk ratio
identical methods apply for a prevalence study involving direct prevalence comparisons
* identical methods apply for a prevalence study involving prevalence odds comparisons
@ the Cox proportional hazards model will yield the same findings, but is more complex to use.

The first step of the initial analysis is to examine the frequency distribution of each variable in order to identify whether the distribution is appropriate for the anticipated statistical model. Since most epidemiological analysis uses categorical variables, it is also appropriate to evaluate how continuous variables may be redefined as categorical variables, for example, dividing age into categories by decade, or dividing an exposure variable into categories representing persons with high, medium, low, or no exposure. The decision as to how to divide a continuous variable into categories usually involves considerations of biological relevance, such as what is a meaningful exposure level, as well as considerations of statistical precision, i.e. establishing category boundaries so that each stratum contains a sufficient number of observations for the statistical analysis.

Initial data analysis may also be undertaken to derive biologically appropriate composite variables based on the individual variables measured during a study. For example, an exposure variable representing average daily exposure to ozone may be calculated based on measurement of the time spent in various locations and the ozone concentration found within each location. (This technique was described in Chapter 4.) Similarly, it is common to estimate daily intake of particular contaminants in food by measuring and combining the intake of all food items known to contain the contaminant under study. As noted in Chapter 5, case definitions for health outcomes may be based on the combination of several separate variables, for example, diagnosis of asthma based on doctor diagnosis or history of wheezing plus indication on lung function tests of reversible airway obstruction or hyperresponsive airways. If the biological interpretation is clear, the derived variables are likely to be used in the subsequent data analysis. Analyzing a few variables rather than a large number of separate, highly related variables is more feasible.

Understanding the relationship between the potentially relevant variables — namely, exposure variables, confounding variables, and effect modifiers including risk factors such as individual susceptibility — is important. The pairwise relationship among the variables (i.e. two variables examined at a time) can be explored visually using a graphical technique, such as a scatterplot of two variables, or an analytical technique such as cross-tabulation for categorical variables. The aim of this analysis is to understand the pattern of relationships among the variables. Relying on tests of statistical significance to decide which variables are "important" or highly correlated would not be appropriate at this stage.

Measures of disease occurrence

Chapter 3 defined the basic measures of disease occurrence that are used in epidemiological studies. This section presents selected statistical analysis methods (using the notation given in Chapter 3 and shown in Table 7.2, which is similar to Table 3.1 on page 45, but with notation for strata). For the reasons outlined above, estimation of confidence intervals is emphasized, but methods for calculating p-values are also presented. Only those data analysis methods that are based on asymptotic theory (i.e. for large samples) are discussed. For analyses of studies involving very small numbers (e.g. if the number of cases of disease is less than 5 in each of the groups to be compared), more complex techniques are required. Readers are referred to standard texts (see, for example, Rothman & Greenland, 1998) for a more comprehensive review.

Table 7.2. Notation for data for stratum in studies of disease incidence and prevalence

Variable	Unit	Exposed	Not exposed	Total
Cases	person	a_i	b_i	M_{1i}
Non-cases	person	c_i	d_i	M_{0i}
Base population	person	N_{1i}	N_{0i}	T_{Ni}
Person-years		Y_{1i}	Y_{0i}	T_{Yi}
Incidence rate		I_{1i}	I_{0i}	
Cumulative incidence		CI_{1i}	CI_{0i}	
Odds		O_{1i}	O_{0i}	

As defined in Chapter 3, the most common measure of disease occurrence in a cohort study is the incidence rate, which is a measure of the disease occurrence per unit time. The incidence rate in the non-exposed group takes the form $I_0 = b/Y_0$ with a standard error (under the Poisson distribution) of:

$$SE(I_0) = \frac{b^{0.5}}{Y_0}$$

and an approximate 95% confidence interval for the incidence rate is thus:

$$I_0 \pm 1.96 \times SE(I_0)$$

The incidence proportion (cumulative incidence or risk) is another measure of disease occurrence in a cohort study that represents the proportion of study subjects who develop the health outcome during a specific period (see pp.46). The incidence proportion in the non-exposed group takes the form $CI_0 = b/N_0$ with a standard error (under the binomial distribution) of:

$$SE(CI_0) = \frac{[b(N_0 - b)]^{0.5}}{N_0^{1.5}}$$

The 95% confidence interval for the incidence proportion is thus:

$$CI_0 \pm 1.96 \times SE(CI_0)$$

Box 7.1. Example of incidence rates

Continuing with the example given in Chapter 3, a study of New Zealand participants in British atmospheric nuclear weapons tests in the Pacific (Pearce et al., 1990) included 1504 men, who served on ships that did not participate in the tests, as the non-exposed group. This group was followed for the period 1957–1987. The 1504 controls contributed a total of 38 937 person-years, and there were 179 deaths. Thus the death rate was 179/38 937 = 46.0 deaths per 10 000 person-years, with a standard error equal to 0.00034 and 95% confidence interval of 39.2 to 52.7 deaths per 10 000 person-years. The incidence proportion of death was 0.119 (11.9%) with a standard error equal to 0.00834, and a 95% confidence interval of 0.103 to 0.135. See Table 7.3 for data related to this example.

As noted in Chapter 3, a cross-sectional study measures the prevalence of a disease or health condition in a population. The measure of disease occurrence in a cross-sectional study is the prevalence proportion ($P_0 = b/N_0$, with b as the number of prevalent cases in a non-exposed population). The statistical properties of the prevalence proportion are similar to those of the

incidence proportion since each is expressed as a proportion of the study population. The standard error and confidence interval for the prevalence proportion are calculated using the same formulae as shown above for the confidence interval.

While most epidemiological studies are based on these measures of disease occurrence, some studies examine health outcomes as a continuous measure. This approach is commonly taken when the outcome is a biological measure, such as blood pressure or the concentration of an enzyme or protein in the blood. (The formulae to calculate a test statistic, standard error, and confidence interval for a normally distributed continuous variable were presented in Section 7.2.)

Table 7.3. Follow-up of New Zealand participants in British atmospheric nuclear weapons tests in the Pacific, 1957–1987

	Test Participants		Controls	
Alive	377	(71.4%)	1085	(72.1%)
Deceased	70	(13.3%)	179	(11.9%)
Emigrated	31	(5.9%)	69	(4.6%)
Unknown	50	(9.5%)	171	(11.4%)
Total	528		1504	
Person-time	13 923		38 937	
Death rate per 10,000 person-years	50.3		46	
SE (death rate)	0.00060		0.00034	
95% CI	38.5,	62.1	39.2,	52.7
Cumulative death risk, 1957-87	0.133		0.119	
SE (death risk)	0.0148		0.0083	
95% CI	0.104,	0.162	0.103,	0.135
Rate ratio	1.09			
Rate ratio chi-square	0.40			
95% CI	0.83,	1.44		
Risk Ratio	1.11			
Risk ratio chi-square	0.67			
95% CI	0.86,	1.44		

Source of data: Pearce et al., 1990.

Measures of effect

The principal measures of effect for data analysis are based on ratios or differences derived from the measures of disease occurrence described above for the various study designs (see Table 7.1).

Cohort studies

The measure of primary interest in a cohort study is often the incidence **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) (see Chapter 3). The hypothesis that the rate ratio is significantly different from the null value (1.0) can be tested with the person-time version of the Mantel-Haenszel chi-square (Mantel & Haenszel, 1959). Essentially this tests whether the observed number of exposed cases differs from the number expected under the null hypothesis that exposure has no effect:

$$\chi^2 = \frac{[Obs(a) - Exp(a)]^2}{Var(Exp(a))} = \frac{[a - Y_1 M_1/T_Y]^2}{[M_1 Y_1 Y_0/T_Y^2]}$$

where M_1, Y_1, Y_0 and T_y are as depicted in Table 7.2. The Mantel-Haenszel chi-square is virtually identical to the usual chi-square test, but is preferable as it is more easily generalized to incorporate multiple strata of confounders (see below). An approximate 95% confidence interval for the rate ratio is then given by (Miettinen, 1976):

$$RR^{1 \pm 1.96/\chi}$$

The second measure of effect used in studies of fixed cohorts is the **risk ratio**. This is the ratio of the proportion of study subjects who develop the health outcome in the exposed group (a/N_1) to that in the non-exposed group (b/N_0) during a specific period. The hypothesis that the risk ratio is significantly different from the null value (1.0) can be tested with the Mantel-Haenszel chi-square (Mantel & Haenszel, 1959):

$$\chi^2 = \frac{[Obs(a) - Exp(a)]^2}{Var(Exp(a))} = \frac{[a - N_1 M_1/T_N]^2}{[M_1 M_0 N_1 N_0/T_N^2(T_N - 1)]}$$

where M_1, M_0, N_1, N_0 and T_N are as depicted in Table 7.2. An approximate 95% confidence interval for the risk ratio is then given by:

$$RR^{1 \pm 1.96/\chi}$$

A third possible effect measure is the incidence **odds ratio**, which is the ratio of the incidence odds in the exposed group (a/c) to that in the non-exposed group (b/d), or as commonly expressed OR = ad/bc. The hypothesis that the odds ratio is significantly different from the null

value (1.0) can be tested with the Mantel-Haenszel chi-square (Mantel & Haenszel, 1959) which has the same form as presented above for the risk ratio.

$$\chi^2 = \frac{[Obs(a) - Exp(a)]^2}{Var(Exp(a))} = \frac{[a - N_1 M_1/T]^2}{[M_1 M_0 N_1 N_0/T^2(T-1)]}$$

An approximate 95% confidence interval for the odds ratio can be estimated using the test-based approach.

$$OR^{1 \pm 1.96/\chi}$$

The reason the same formula can be used for the risk ratio and the odds ratio is that both measures are based on proportions which, therefore, can be analyzed using the binomial distribution approach described in Section 7.2.

An analogous approach can be used to calculate measures of effect based on differences rather than ratios, in particular the rate difference and the risk difference. The **rate difference** takes the form $RD = I_1 - I_0 = a/Y_1 - b/Y_0$. The hypothesis that the rate difference differs significantly from the null value (0.0) can be tested with the person-time version of the Mantel-Haenszel chi-square (Mantel & Haenszel, 1959) which has the same form as presented above in the context of the rate ratio. An approximate 95% confidence interval for the rate difference is given by:

$$RD(1 \pm 1.96/\chi)$$

The **risk difference** takes the form $RD = CI_1 - CI_0 = a/N_1 - b/N_0$. The hypothesis that the risk difference is significantly different from the null value (0.0) can be tested with the Mantel-Haenszel chi-square (Mantel & Haenszel, 1959) which has the same form as presented above in the context of the risk ratio. An approximate 95% confidence interval for the risk difference is calculated using the same formula as shown above for the rate difference.

Case–control studies

In case–control studies, the relative risk measure is the exposure **odds ratio** which is the ratio of the exposure odds (a/b) in the case group to that in the control group (c/d). The odds ratio takes the form OR = ad/bc, which is mathematically the same formula as for the incidence odds ratio in a cohort study. Therefore, the same analytical methods can be use as those presented above for the incidence odds ratio. The hypothesis that the odds ratio is differs significantly from the null value (1.0) can be tested with the Mantel-Haenszel chi-square (Mantel and Haenszel, 1959). An approximate 95% confidence interval for the odds ratio can be estimated using the test-based approach as shown above.

Table 7.4 presents findings from a case–control study by Sunyer et al. (1989) of serum immunoglobulin-E antibodies reactive with soybean in epidemic asthma. The cases constituted 86 patients who had had an emergency room admission for asthma during asthma epidemics in Barcelona during 1987; 86 controls were selected from emergency room admissions for asthma during non-epidemic periods. Serum from 64 of the cases — but only 4 of the controls — was

reactive with commercially prepared soybean antigen, yielding a very high odds ratio of 59.6 (95% CI 25.3–140.8). The study thus revealed a strong association between serum IgE antibodies and commercially prepared soybean antigen. It thus complemented other work by the same group which found that the epidemics had occurred only on days when soybeans were unloaded at the city's port (Antó et al., 1989).

Table 7.4. Serum reactivity for commercially prepared soybean antigen and epidemic asthma in Barcelona

Soybean antigen	Cases	Controls	Odds Ratio	95% CI
Reactive	64	4	59.6	25.3–140.8
Not reactive	22	82		

Source: derived from Sunyer et al., 1989.

Prevalence studies

As discussed in Chapter 3, two relative risk estimates can be calculated based on prevalence studies. The **prevalence ratio** is the ratio of the prevalence proportion for the exposed to the prevalence proportion for the unexposed. Such estimates can be used, for example, when investigating an epidemic outbreak, in which case the prevalence ratio provides an estimate of the risk ratio. The statistical methods for estimation and testing of the prevalence ratio are identical to those used to calculate the risk ratio in a cohort study.

The **prevalence odds ratio** is a second effect measure used in prevalence studies. Assuming that the study population remains in a "steady state" over time and that exposure and disease status are unrelated to immigration and emigration rates, then the prevalence odds are equal to the incidence rate multiplied by the average disease duration $\left(p(1-p) = I\overline{D}\right)$. Thus, if exposure does not affect the mean duration of disease, then the prevalence odds ratio ($OR = I_1D_1/I_0D_0$) directly estimates the incidence rate ratio. In a prevalence study of chronic health conditions, the prevalence odds ratio is therefore a common effect measure. The statistical methods for estimation and testing of the prevalence odds ratio are identical to those for odds ratios in other contexts (see above.)

Table 3.2 on page 50 presented the findings of a prevalence study of respiratory symptoms and exposure to automobile exhaust in 1517 people (Nitta et al., 1993). Of these, 570 lived within 20m of the roadside and of these 56 (9.8%) reported chronic wheezing. The other 578 participants lived 20–150m from the roadside and 24 (4.2%) of these reported chronic wheezing. Thus, the odds of chronic wheezing of those living close to the roadside was 56/514, the odds in those living further away was 24/554, and the odds ratio was 2.5 (95% CI 1.6–4.1).

Dose–response relationships

Analysis of a dose–response relationship can be undertaken by comparing sub-groups of the study population with different exposure levels, using the previously described analytical methods. If the exposure information is based on an ordinal scale — for example, if study participants can be classified as experiencing high, medium, low or no exposure — the non-exposed group can be taken as the reference group, and each of the three exposure groups compared with this group.

Table 7.5 gives details of a prevalence case–control study conducted by Flatt et al. (1990) of 56 asthmatic patients (cases) and 59 non-asthmatic controls in Dunedin, New Zealand. In this region, environmental selenium concentrations are low, with the result that dietary intake of selenium is also low. Since selenium has antioxidant properties, the investigators hypothesized that selenium deficiency may play a role in the development of asthma in this region. The asthma cases were identified through an outpatient chest clinic and a general practice. Controls were selected from a general medical outpatient and blood donor population, and from the families of the asthmatic cases. A 10 ml venous blood sample was collected from each subject for assay of glutathione peroxidase activity in whole blood as a biological indicator of selenium deficiency. (Selenium is an essential trace metal and crucial to the production of glutathione peroxidase.) The findings showed a 5.8-fold increased risk of asthma in subjects whose whole blood glutathione peroxidase activity fell with the lowest range of those sampled. The investigators concluded that these findings were consistent with the hypothesis that low selenium concentration may have a role in the pathogenesis of asthma in New Zealand.

Table 7.5. Findings from an asthma prevalence case–control study

Whole blood glutathione activity (units/g Hb)	Cases	Controls	Odds ratio	95% CI
≥ 30	7	13	1	–
25–29	14	19	1.4	0.4–4.5
20–24	15	20	1.6	0.5–5.2
< 20	20	7	5.8	1.6–21.2

Source: Flatt et al., 1990.

If continuous exposure information is available (as in the above example), the exposure is usually first categorized (e.g. high, medium or low) and a preliminary analysis conducted, using the methods outlined above. This approach is advisable, not only to avoid introducing errors through use of complex mathematical models, but also to enable investigation of the possible shape of the exposure–response curve. Any further analysis involving continuous data will necessitate the use of regression methods. These are discussed in Section 7.5.

7.4 Stratified analysis

Stratification for data analysis means that the observations are divided up into strata according to one or more variables other than that of exposure or the disease outcome (Ahlbom, 1993). For example, subjects may be divided into strata based on age or gender in a study of the relationship between air pollution exposure and respiratory disease. As described in Chapter 6, stratification can be performed when selecting individuals for a study. The individuals are then distributed across the strata according to the proportions determined in advance. Stratified analysis is routinely used both in conjunction with a stratified study design and even if subjects were not stratified at the time of recruitment into the study.

A stratified analysis is performed for two different reasons: to control confounding and to facilitate the assessment of effect modification. Since confounding refers to a risk factor that is distributed differently among exposed and unexposed, a possible consequence of not controlling for confounding would be a distortion in the effect measure since the exposed and unexposed groups would differ in disease occurrence independently of any effect of the studied exposure. For example, in a cohort study of the effect of cigarette smoking on cardiovascular disease, stratifying on age would be appropriate since age is a risk factor for cardiovascular disease and the proportion of persons who smoke may vary among different age groups. If older persons, who have a higher cardiovascular disease risk, tend to smoke less than younger persons, an unstratified or "crude" analysis would underestimate the true relative risk.

Stratified analysis is also used to evaluate any effect modification, which occurs when the effect of exposure depends on the level of another factor in the source population (Kleinbaum et al., 1982; see pp.52). Consequently, the effect measure would vary among the different strata. Stratified analysis can be used to estimate the effect measure separately for single strata. Effect modification is evaluated by comparing strata.

Analysis for confounding

Control of confounding requires that the data are stratified into sub-groups according to the levels of the confounder(s), and that a summary effect is calculated which summarizes the information across strata. For example, controlling for age, using five categories, and for gender, using two categories, might involve grouping the data into the 10 (= 5 x 2) confounder strata and calculating a summary rate ratio which is a weighted average of the stratum-specific rate ratios. Two approaches are available for calculating a summary effect measure: **pooling** and **standardization**.

Pooling

Pooling is the most common approach to stratified analysis. It consists of calculating a summary effect estimate, based on the assumption that the stratum-specific exposure effects are uniform and, accordingly, that any observed variation among the stratum-specific effect estimates is due to random error. Weights are used to give greater emphasis to strata with greater precision. The system of weights originally proposed by Mantel and Haenszel (1959) is the most widely used method for obtaining a summary effect estimate. This system of weights for pooling in stratified analysis is relatively simple and close to being statistically optimal (even if the numbers in some

strata are small). Mantel-Haenszel methods have now been developed for stratified analysis for each of the study designs and effect measures described above.

The Mantel-Haenszel summary **rate ratio** uses weights of:

$$w_i = \frac{b_i Y_{1i}}{T_{Yi}}$$

and takes the form:

$$RR = \frac{\Sigma w_i RR_i}{\Sigma w_i} = \frac{\Sigma a_i Y_{0i}/T_{Yi}}{\Sigma b_i Y_{1i}/T_{Yi}}$$

The hypothesis that the summary rate ratio differs significantly from the null value (1.0) can be tested with the person-time version of the Mantel-Haenszel summary chi-square (Mantel & Haenszel, 1959):

$$\chi^2 = \frac{[\Sigma Obs(a) - \Sigma Exp(a)]^2}{\Sigma Var\ (Exp(a))} = \frac{[\Sigma a_i - \Sigma Y_{1i} M_{1i}/T_{Yi}]^2}{[\Sigma M_{1i} Y_{1i} Y_{0i}/T_{Yi}^2]}$$

where M_{1i}, Y_{1i}, Y_{0i} and T_{Yi} are as depicted in Table 7.2. An approximate 95% confidence interval for the summary rate ratio is then given by:

$$RR^{1 \pm 1.96/\chi}$$

The Mantel-Haenszel summary **risk ratio** uses weights of:

$$w_i = \frac{b_i N_{1i}}{T_{Ni}}$$

and takes the form:

$$RR = \frac{\Sigma w_i RR_i}{\Sigma w_i} = \frac{\Sigma a_i N_{0i}/T_{Ni}}{\Sigma b_i N_{1i}/T_{Ni}}$$

The hypothesis that the summary risk ratio differs significantly from the null value (1.0) can be tested with the Mantel-Haenszel summary chi-square (Mantel & Haenszel, 1959):

$$\chi^2 = \frac{[\Sigma Obs(a) - \Sigma Exp(a)]^2}{\Sigma Var(Exp(a))} = \frac{[\Sigma a_i - \Sigma N_{1i} M_{1i}/T_{Ni}]^2}{[\Sigma M_{1i} M_{0i} N_{1i} N_{0i}/T_{Ni}^2]}$$

where M_{1i}, M_{0i}, N_{1i}, N_{0i} and T_{Ni} are as depicted in Table 7.2. An approximate 95% confidence interval for the summary cumulative incidence ratio is then given by:

$$RR^{1 \pm 1.96/\chi}$$

The Mantel-Haenszel summary **odds ratio** uses weights of:

$$w_i = \frac{b_i N_{1i}}{T_i}$$

and takes the form:

$$OR = \frac{\Sigma w_i OR_i}{\Sigma w_i} = \frac{\Sigma a_i d_i/T_i}{\Sigma b_i c_i/T_i}$$

The hypothesis that the summary odds ratio differs significantly from the null value (1.0) can be tested with the Mantel-Haenszel summary chi-square (Mantel & Haenszel, 1959):

$$\chi^2 = \frac{[\Sigma Obs(a) - \Sigma Exp(a)]^2}{\Sigma Var(Exp(a))} = \frac{[\Sigma a_i - \Sigma N_{1i} M_{1i}/T_i]^2}{[\Sigma M_{1i} M_{0i} N_{1i} N_{0i}/T_i^2]}$$

An approximate 95% confidence interval for the summary odds ratio is then given by:

$$OR^{1 \pm 1.96/\chi}$$

Box 7.1. Example of stratified analysis for case-control study

This box presents data from an occupational epidemiology study of cancer of the nasal cavity and paranasal sinuses and formaldehyde exposure (Hayes et al., 1986), stratified according to the confounding factor of wood dust exposure.

Exposure of cases and controls in a study of cancer of the paranasal sinuses and formaldehyde exposure

Low wood dust exposure		
Formaldehyde	Cases	Controls
Exposed	15	18
Not exposed	48	143
Total	63	161

High wood dust exposure		
Formaldehyde	Cases	Controls
Exposed	16	16
Not exposed	12	18
Total	28	34

The overall odds ratio, adjusted for wood dust exposure, is:

$$OR = \frac{(15\times143/224) + (16\times18/62)}{(48\times18/224) + (12\times16/62)} = 2.05$$

and the Mantel-Haenszel chi-square is:

$$\chi^2_{MH} = \frac{[31 - 23.73]^2}{9.61} = 5.50$$

The square-root of 5.50 is 2.35, and the test-based 95% confidence limits are thus:

$$OR_L = 2.05^{1-1.96/2.35} = 1.13$$
$$OR_U = 2.05^{1+1.96/2.35} = 3.73$$

Standardization

Confounding can also be controlled through a stratified analysis by calculating a **standardized rate** for each of the groups to be compared. This approach involves calculating a weighted average of the stratum-specific rates and then calculating the ratio of the standardized rates. Note that both the pooling and standardization approaches calculate a summary effect estimate; however, the approaches differ in two ways. First, in pooling, stratum-specific rate ratios are calculated and then a weighted average of the ratios is derived, while in standardization, a weighted average of the rates in the different sub-groups is calculated and then the ratio of the weighted rates derived. Second, in calculating a standardized rate, no assumption is made

regarding whether stratum-specific effects are uniform. The formulae for the weights therefore differ between the two approaches. The general form of a standardized rate is:

$$SR = \frac{\Sigma\, w_i\, I_i}{\Sigma\, w_i}$$

where I_{1i} is the incidence rate in stratum I. If SR_1 is the standardized rate in the exposed group and SR_0 is the standardized rate in the comparison group (Table 7.1), the **standardized rate ratio** is:

$$SRR = \frac{SR_1}{SR_0}$$

The weights can be selected somewhat arbitrarily. However, two main alternatives have become widely used when choosing the weights since they lead to a clear interpretation of the standardized effect measures. First, the weights can be taken from the confounder distribution of the comparison population (i.e. $w_i = Y_{0i}$ — which represents person-time distribution by strata of the confounding variable in the comparison population). In this instance the standardized rate for the comparison population (SR_0) is the same as the crude rate for this population. Using this approach, which is one type of **direct standardization**, the standardized rate ratio is thus:

$$SRR = \frac{\Sigma\, Y_{0i}(a_i/Y_{1i})}{\Sigma b_i}$$

The SRR is therefore the ratio of the expected number of cases in the comparison population (based on the rates in the exposed group) to the number of observed cases in the comparison population.

An alternative approach is **indirect standardization** in which the weights are taken from the confounder distribution of the exposed population (i.e. $w_i = Y_{1i}$). The standardized rate for the exposed population (SR_1) is thus the same as the crude rate for this population, and the standardized rate ratio becomes:

$$SMR = \frac{\Sigma a_i}{\Sigma Y_{1i}(b_i/Y_{0i})}$$

The standardized mortality (or morbidity) ratio (SMR) is therefore the ratio of the number of observed cases in the exposed population to the expected number of cases in the exposed population, based on the rates in the comparison group.

The SMR is primarily used in cohort studies in which the death rate (or incidence rate) in a particular population, such as a factory workforce or a community, is compared with national death rates. In other words, the observed number of deaths in the community is compared with the number of deaths that would be expected on the basis of national death rates. If national death rates are used as the comparison, the SMR has the advantage of statistical stability because the rates used for the comparison population are derived from the entire national population. However, when more than two groups are being compared, the SMRs are not directly comparable, since they are based on different weights. Conversely, the SRR does have the advantage of comparability across groups since the confounder distribution of the reference group serves as the weight in each instance. The SRR may provide unstable estimates, though, if the observed numbers are small.

Analysis of a matched study design

Matching is rarely performed for cohort studies because typically the cost of examining potential subjects to determine their status on matching variables (other than age, gender, or family relationship) makes this approach inefficient. If matching is undertaken in a cohort study, the same approaches and formulae described above for unmatched, stratified data can be used. A matched study design is used more often for case–control studies than for cohort studies. However, frequency matching is used more often than individual matching. Data for a frequency-matched case–control study can be analyzed using the approaches and formulae described above matched, stratified data. Different formulae which take into account the matching study design must be used for analysis of individually-matched case–control studies. Standard textbooks can be consulted for more complicated matching study designs (e.g. multiple matched designs and variable matching ratio designs) (Kleinbaum et al., 1982; Rothman & Greenland, 1998).

The data layout and notation for analyzing an individually matched case-control study are shown in Table 7.6. Each cell in the table represents the number of **pairs** of case and control-matched subjects who have the exposure characteristic. For example, the "f" cell is the number of case–control pairs in which both the case and the control are exposed, while the "g" cell is the number of pairs in which the case subject is exposed but the control subject is not exposed.

Table 7.6. Notation for data in a matched case–control study

			Controls	
			Exposed	
			+	-
Cases	Exposed	+	f	g
		-	h	j

The odds ratio for a matched study design controls for confounding due to the matching variable. The matched pair odds ratio takes the form:

$$OR_{MP} = \frac{g}{h}$$

It may be noted that only subject pairs which are discordant with respect to exposure contribute information to the calculation of the odds ratio. Pairs in which both case and control are exposed or in which both are unexposed do not contribute information to calculation of the odds ratio.

The chi-square for a matched study design is:

$$\chi^2 = \frac{(g-h)^2}{g+h}$$

where g and h are as shown in Table 7.6. An approximate 95% confidence interval for the matched pair odds ratio is then given by:

$$OR^{1 \pm 1.96/\chi}$$

Analysis of effect modification

Determining whether the effect of exposure varies among the different sub-groups of the study base population is commonly of interest. For example, the investigators may wish to examine whether the effect of air pollution on respiratory symptoms varies between men and women. Such **effect modification** is also assessed by stratification, but is distinct from confounding in that it does not represent a bias (which should be removed or controlled), but rather a real difference in the effect of exposure on the various sub-groups. Effect modification can be evaluated by calculating the effect measures (e.g. rate ratio) separately for each stratum. The stratum-specific effect measures can then be compared to evaluate whether the effect of exposure varies among the strata. Calculating a summary effect measure is not appropriate if analysis of the separate strata has already provided evidence of effect modification.

Table 3.3 on page 53 presented data from a study of asbestos exposure, tobacco smoking and lung cancer (Hammond et al., 1979), which provides an example of effect modification. If the rate difference is applied to asbestos exposure, 470 per 100 000 person-years are at risk among smokers and 47 per 100 000 person-years among non-smokers. It is clearly evident that smoking modifies the effect of asbestos exposure. However, if the rate ratio is used, the relative risk from asbestos exposure is about five in both smokers and non-smokers. Asbestos exposure therefore appears to multiply the lung cancer death rate by about five times, and this effect is not modified by smoking if a multiplicative effect measure (such as the rate ratio) is used. However, smoking does appear to modify the effect of asbestos exposure if an additive effect measure (such as the rate difference) is used. In other words, there is no interaction between smoking and asbestos exposure on a multiplicative scale, but there is an interaction on an additive scale (Pearce, 1989).

7.5 Multivariate analysis

Usually it is not possible to control for more than two or three confounders in a stratified analysis. Finer stratification tends to lead to too many strata containing no exposed or no unexposed persons (or no cases or non-cases). Such strata are uninformative, and hence fine stratification is wasteful of information. However, the problem of multiple confounders can be countered to some extent through use of multivariate methods which allow the simultaneous control of a greater number of confounders through "smoothing" of the data across confounder strata. This section examines some of the multivariate methods which can be used for analyses of environmental epidemiology studies (see e.g. Kleinbaum et al., 1982; Kleinbaum et al., 1988; Rothman & Greenland, 1998). These analyses can be performed with standard statistical packages such as GLIM (Baker & Nelder, 1978), BMDP (BMDP, 1979), SAS (SAS, 1988) and Egret (SERC, Seattle, WA), but familiarity with statistical concepts and statistical software packages is crucial. Such analyses are therefore best performed by or in collaboration with a biostatistician. The concepts underlying multivariate analysis are described below.

Linear regression

The most common form of regression is linear, whereby the mathematical relationship between variables is represented by a straight line. One of the variables is considered to be the dependent variable and the degree to which its value varies, according to one or more independent variables, is examined. Linear regression can be applied if the health outcome that is being studied is continuous. Standard regression methods for normally distributed data can be used. The relevant statistical techniques are explained clearly in many standard texts (see, for example, Kleinbaum et al. , 1988). Linear regression may also be appropriate for analysis of ecological study data in which exposure measures and health outcome rates are compared among multiple populations (ecological units), although the regression techniques may be more complicated because of the need to weight the observations by their precision. (Precision is generally related to the size of the ecological units).

Log-linear models

In most environmental epidemiology studies the health outcome that is being investigated is dichotomous: i.e. the study participants are classified as to whether or not they experience a particular health outcome (e.g. lung cancer, cardiovascular disease). The dichotomous health outcome variable is converted to a rate or a risk as discussed above.

The customary approach to modeling the relationship between health outcome and risk factors is to calculate the log of the disease rate (or risk or odds) as a linear combination of a set of risk factors, comprising the main exposure(s) of interest and the potential confounders (Pearce et al., 1988). The model takes the general form:

$$\ln(Y) = b_0 + b_1X_1 + b_2X_2 + \ldots + b_jX_j + e$$

Where ln(Y) is the natural logarithm of the outcome Y, b_i the coefficients of the risk factors, and X_i the various risk factors. Note that if all of the X_i's are zero then $\ln(Y) = b_0$. Thus $Y_0 = \exp(b_0)$ estimates the disease risk (or rate or odds) in the group of persons with zero values for each of the X_i's. (The random error term "e" is omitted from subsequent equations.)

The coefficient estimates for the models are obtained by the method of **maximum likelihood**. (Maximum likelihood methods can also be used in a stratified analysis, but the computations are very complex.) This method is based on the **likelihood function** which represents the probability of observing the data as a function of the unknown parameters (b_0, b_1 ... b_j). The method involves finding the values of the unknown parameters which maximize the value of the likelihood function.

The example below is of a simple situation in which exposure is dichotomous and represented by X_1 ($X_1 = 1$ if exposed, $X_1 = 0$ if non-exposed), and the only confounder is age, which is stratified into two levels and represented by X_2 (e.g. $X_2 = 1$ if age ≥ 55, $X_2 = 0$ if age < 55). If a particular age-group is considered, the model for the exposed sub-group ($X_1 = 1$) is:

$$\ln(Y_E) = b_0 + b_1 + b_2X_2$$

and the model for the unexposed group ($X_1 = 0$) is:

$$\ln(Y_{\bar{E}}) = b_0 + b_2X_2$$

Subtracting the second equation from the first yields:

$$\ln(Y_E / Y_{\bar{E}}) = b_1$$

and "exponentiating" both sides yields the relative risk in exposed persons compared to non-exposed persons:

$$R = \exp(b_1)$$

The coefficient b_1 associated with the exposure term (X_1) can thus be used to calculate the relative risk for exposed persons compared to non-exposed persons.

Modifying the log-linear model to allow for multiple levels of exposure is straightforward. Just as one variable is needed to denote a factor with two levels, k-1 variables are needed to denote a factor with k levels. A person in the reference category is assigned scores of zero for each of the k-1 indicator variables. The coefficients for each of the exposure categories, when "exponentiated", estimate the relative risks for the various categories compared to the reference category.

The log-linear model does not distinguish between the main exposure(s) and confounders; the exposures and confounders are all modeled as "risk factors", and the coefficient for any factor estimates its effect on the health outcome, adjusted for the effects of all other factors in the model. Construction of indicator variables for confounders with multiple levels is therefore similar to that undertaken for the main exposure. For example, four indicator variables would be used to denote five age-groups.

It is preferable to use a categorical definition of the exposure variable (Rothman & Greenland, 1998) as the use of continuous exposure variables can be problematic. a model that uses continuous exposure variables assumes that exposure is exponentially related to disease risk (Greenland, 1979). Thus each additional unit of exposure *multiplies* the relative risk by a constant value. For example, suppose that the coefficient for a particular factor is 0.693. Exponentiating this term yields a relative risk of 2.0 for one unit of exposure, whereas a person with two units of exposure (X_1=2) has an estimated relative risk of exp(2x0.693) = 4.0 (=2.0^2), compared to a person with no exposure. Such an exponential exposure–response relationship is rarely observed in epidemiological data and a better model is surely needed. Using a categorical exposure classification has the advantage that no assumptions are made about the shape of the exposure–response relationship, and relationships which do not fit an exponential pattern can be detected.

An objection that is often raised to the categorical approach is that continuous exposure information and statistical power are lost, and several terms are required in the model rather than just one. Statistical power is indeed reduced, relative to the continuous approach, if the exposure–response relationship follows a smooth monotonic pattern throughout the exposure range. Nevertheless, statistical power is a secondary consideration in epidemiological studies. The main goal is to obtain valid estimates of the actual exposure–response relationship. This is achieved most easily by means of a categorical analysis.

As for the main exposure, use of a set of categorical variables to denote the levels of the confounders is to be favored. Use of continuous confounder variables is most appropriate when the relationship with disease is known to be approximately exponential. For example, an exponential relationship with age does appear to hold approximately for certain diseases or health outcomes, such as solid tumors. However, before using continuous variables, categorical analysis should be performed, to verify that the data do show an exponential relationship (or at least one that is increasing monotonically).

When estimating the exposure effect adjustment for all potential confounders is usually desirable. However, if a strong correlation exists between some risk factors, the model will be unstable due to multicollinearity. Elimination of some of the highly correlated potential confounders from the model (see below) may then be necessary.

Specific applications

Table 7.1 summarizes the multivariate methods appropriate for the analysis of various types of epidemiological study. These analyses can be performed using standard statistical packages such as EPICURE or STATA.

Poisson regression is an extension of simple incidence rate analyses involving person-time denominators and is thus the appropriate method of analysis for cohort studies involving person-time data (Table 7.1). The Poisson regression model for a rate ratio has the general form:

$$\ln(\lambda) = b_0 + b_1X_1 + b_2X_2 + \ldots + b_jX_j$$

where λ is the incidence rate for persons with specified values of X_1, X_2, ... X_j. Poisson regression thus directly follows on from simpler methods that involve generating person-time data, calculating incidence rates in each confounder strata (e.g. each age-group) and obtaining overall rate ratios adjusted for confounders (e.g. by the Mantel-Haenszel method). For Poisson regression analysis, as for stratified analysis of rate ratios, it is first necessary to calculate the person-years in each exposure and confounder category in a time-related manner (Pearce and Checkoway, 1987). It is also possible to use *binomial regression* to estimate risk ratios instead of rate ratios in a cohort study.

The corresponding approach for analyzing case–control data is the *logistic regression* model, where Y is the odds of a person being a case (= P/(1 - P), and P the proportion of persons who are cases). The effect measure of interest is the odds ratio. The logistic regression equation takes the form:

$$\ln(P/(1-P)) = b_0 + b_1X_1 + b_2X_2 + \ldots + b_jX_j$$

where P is the proportion of persons who are cases in the group of persons who possess characteristics X_1, X_2 ... X_j. Logistic regression is thus the modeling analogue of the standard Mantel-Haenszel procedure (Mantel & Haenszel, 1959).

Table 7.7 presents the results of a logistic regression analysis of a case–control study of congenital malformations and maternal drinking-water supply (Dorsch et al., 1984). Each of the variables was entered into the model simultaneously (together with a number of other potential confounders). The table thus shows the independent effect of each factor, adjusted for the effects of all other potential risk factors for which information was available. The table indicates that, in comparison with women who drank only rainwater, women who drank well-water or water from Blue Lake were at increased risk of bearing a malformed child. The authors concluded that these findings were consistent with the hypothesis of nitrate-induced teratogenesis.

Conditional logistic regression is the procedure used for multivariate analysis of matched study data. It is based on the likelihood of the data being distributed into different strata, as they are, given the marginal totals in each confounder stratum. Once matching of a number of factors has been conducted and genuine "pairs" found to exist (e.g. twins), then a fully matched analysis, using conditional logistic regression, is appropriate. However, if matching has been performed only in relation to general factors such as age and sex, an individually matched analysis and a stratified analysis will yield equally valid findings, although the latter will provide a more precise effect estimate. Hence, the more straightforward unconditional estimation procedure is generally used when the strata are large.

Table 7.9. Results of a logistic regression analysis of risk of malformations in relation to maternal water supply in the Lower South East of South Australia, 1951–1979

Variables	beta	SE	Odds ratio	95% CI
Baby's sex				
female			1.0	
male	0.48	0.2	1.6	1.1–2.4
Mother's marital status				
married			1.0	
single	-0.12	0.54	0.9	0.3–2.6
other	0.15	0.68	1.2	0.3–4.4
Mother's nationality				
Australian			1.0	
British	0.16	0.43	1.2	0.5–2.7
European	-0.31	0.38	0.7	0.4–4.5
Other	-1.16	0.93	0.3	0.1–1.9
Father's employment				
Class 1			1.0	
Class 2	0.13	0.34	1.1	0.6–2.3
Class 3	-0.23	0.40	0.8	0.4–1.7
Class 4	0.13	0.41	1.1	0.5–2.6
Mother's area of residence				
Mt. Gambier — urban			1.0	
Mt. Gambier — rural	1.23	0.46	3.4	1.4–8.5
Other urban	0.79	0.51	2.2	0.8–5.9
Other rural	0.59	0.48	1.8	0.7–4.7
Mother's water supply				
Rainwater			1.0	
Blue Lake	1.59	0.45	4.9	2.1–11.7
Bore	1.46	0.59	4.3	1.4–13.8
Other mains	0.38	0.55	1.5	0.5–4.3
Not known	1.53	0.47	4.6	1.8–11.6

Source: Dorsch et al., 1984.

Modeling issues

Multivariate methods have become increasingly popular in recent years. These are powerful statistical techniques that have many benefits when used appropriately and in combination with more straightforward methods. However, they are frequently used inappropriately, and strong modeling assumptions are necessary. It is always important to check these assumptions, to try alternative model forms, and to check the results against those of simple stratified analyses. These simpler analyses should be performed first. The modeling strategy should then be specified in advance, so that only those models of *a priori* interest are assessed, in order to avoid

"data dredging", in which investigators run hundreds of models and then select the findings which they most like.

It needs to be stressed that statisticians and epidemiologists may use multivariate models differently from one another. In conventional statistical modeling, the emphasis is on prediction. The aim is to achieve a model which "fits well", with the minimum number of variables. So in deciding whether or not to include particular variables in a model, the principal considerations are statistical significance and goodness-of-fit. Standard stepwise regression methods have been developed and are mostly, but not always, effective in achieving these goals.

Epidemiologists, on the other hand, use models for effect estimation and etiological understanding. In most epidemiological studies, one principal exposure and several potential confounders are being investigated. The main exposure variable should be included in all models irrespective of whether it is statistically significant. The decision as to whether to add potential confounders to the models will depend on whether the main exposure effect estimate changes (and not on whether the term for the confounder is statistically significant). Potential confounders should nearly always be controlled for, but problems may arise if they are closely correlated with exposure, since multicollinearity may then occur, rendering the model unstable. Therefore, the principle concerns here are controlling confounding while avoiding multicollinearity.

Multicollinearity occurs when a variable is nearly a linear combination of the other variables in the model. In particular, if there is a strong correlation between a confounder(s) and the main exposure, the exposure effect estimate will be unstable and its standard error large. If deleting a confounder(s) leaves the main effect estimate virtually unchanged, but greatly reduces its standard error, this will suggest that the "smaller" model provides more precise estimates of the main effect without compromising validity. However, it is less clear how to proceed when a strong confounder is a source of multicollinearity, since the increase in precision due to deleting such a confounder may be offset by an increase in bias due to inadequate control of confounding (Robins & Greenland, 1986).

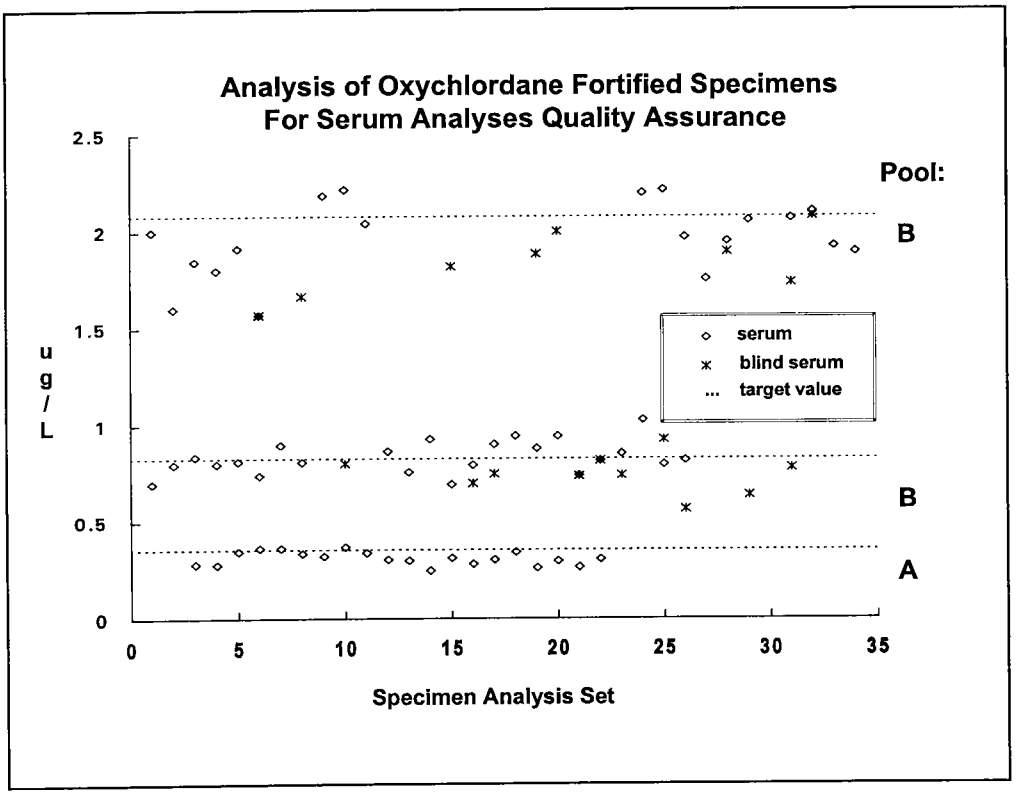
Influential data points are data points that influence the results of the regression analysis unduly. This problem can be largely avoided by using only categorical variables. This applies particularly if fine categorization is used to isolate extreme data points in separate categories, from which they will automatically be discarded if no comparison data are available (Rothman & Greenland, 1998). Influential data points are of more concern in analyses that rely on continuous variables. An assessment procedure can be carried out which involves deleting each data point in turn in order to ascertain whether inclusion of that data point affects the main effect estimate "considerably".

Most tests for **goodness of fit** involve grouping the data and comparing the observed number of cases in each group with that predicted by the model (usually with a chi-square statistic). Although the value of goodness of fit tests is evident in the statistical context, where prediction is often a primary goal, this is not so in the epidemiological context. For example, a poorly fitting model may still produce valid effect estimates. The lack of fit indicates the existence of a strong risk factor, or an interaction of risk factors already included in the model, which has not been adequately controlled. But it does not indicate whether the uncontrolled factor is associated with exposure and hence is a confounder (Checkoway et al., 1989).

7.6 Treatment and analysis of misclassification

The most effective strategy for reducing bias due to measurement error is quite simply to pay meticulous attention to this issue when designing a study and collecting the data. This means that clear criteria and procedures should be established for data collection, and study personnel trained and monitored throughout the study. Quality assurance/quality control procedures should also be defined at the beginning of the study to enable monitoring of data collection, laboratory analysis, and data management. Additionally, quality assurance data can be analyzed during a study to ensure that data collection and laboratory analysis techniques are accurate. For example, in a population-based study of organochloride pesticide concentration in human serum, investigators included a quality-assurance specimen with known concentrations of pesticides (made from pooled sera) in each batch of specimens to be analyzed. Using the quality-assurance specimens, the investigators monitored the validity of laboratory analyses to verify that the results were consistent from batch to batch over time (Fig. 7.1) (Baker et al., 1991; Wolff et al., 1991).

Fig. 7.1. Example of a quality assurance analysis for laboratory analysis of serum pesticide concentrations



Source: Wolff, et al., 1991.

During the analysis of a study, investigators typically examine quality assurance/quality control data to evaluate whether any systematic errors occurred during data collection, and to estimate data reliability. This information is taken into consideration when interpreting the study findings, especially when considering the possibility of misclassification bias.

The effect of using inaccurate exposure variables

Determining the quantitative effects of measurement error on observed study results can be quite complicated because errors can be either differential or non-differential, and can affect measurement of both exposure and disease outcome variables (as well as confounders and effect modifiers). To simplify this discussion, the effects of measurement error are presented for study results for which measurement of disease outcome is accurate and for which error occurs with respect to non-differential exposure measurement only.

As discussed in Chapter 3 in connection with non-differential measurement error, the observed measure of effect is biased toward the "null". The amount of bias is a function of the true effect measure and the precision of the exposure measure. a measure of precision is the validity coefficient, ρ_{tx}, which is the correlation of the true measure, T, with the observed measure, X. Thus ρ_{tx}^2 is the proportion of variance of X explained by T. For a continuous outcome measure and continuous exposure measure, the observed regression coefficient is:

$$\beta_O = \rho_{TX}^2 \beta_T$$

The effect of error in a continuous exposure measure on the odds ratio is:

$$OR_O = OR_T^{\rho_{TX}^2}$$

For example, suppose a study was undertaken on the association between lifetime exposure to the pesticide *p,p'*-DDT and breast cancer incidence. If pesticide exposure doubles the risk of breast cancer, then OR_T would equal 2.0. If the correlation between the estimated pesticide exposure and true exposure is 0.6, then the observed odds ratio would be $2.0^{0.36} = 1.28$. This example shows that non-differential measurement error can result in substantial bias in the effect estimates.

The validity of dichotomous measures of exposure is commonly expressed in terms of sensitivity and specificity. The **exposure sensitivity** represents the proportion of those truly exposed who have been correctly classified as being exposed by the exposure variable used in the study. The **exposure specificity** represents the proportion of those who have been correctly classified as unexposed. The observable proportions (p) of those exposed in the diseased and non-diseased groups are related to the true proportions (P) of those exposed, and can be expressed as follows (Armstrong et al., 1992):

$$p = sensitivity * P + (1 - specificity) * (1 - P)$$

If the degree of exposure misclassification is equal in the diseased and the control subjects (i.e. when misclassification is **non-differential** so that the sensitivity and specificity for the disease group are equal to those for the control group), the observable odds ratio is, by definition, equal to:

$$OR_O = [p_D * (1 - p_N)] / [p_N * (1 - p_N)]$$

Table 7.8 presents a number of numerical results for different odds ratios and different values of sensitivity and specificity in order to illustrate the magnitude of bias that can introduced by measurement error.

Table 7.8. Bias in observed odds ratios as a function of sensitivity and specificity of exposure measurement method *

Exposure sensitivity	Exposure specificity	True odds ratio	
		2	4
		Observed odds ratio	
0.6	0.8	1.29	1.71
0.8	0.8	1.42	2.13
0.95	0.8	1.52	2.51
0.6	0.95	1.6	2.53
0.8	0.95	1.71	2.96
0.95	0.95	1.8	3.35
0.6	0.99	1.77	2.99
0.8	0.99	1.86	3.37
0.95	0.99	1.94	3.74
0.99	0.99	1.96	3.85

* This example assumes that 0.2 (20%) of the control group are exposed.

Adjustment for inaccurate exposure variables

Some epidemiologists recommend using adjustment procedures to estimate quantitatively the amount of bias in study results. These procedures typically use correlation measures derived from validation studies to adjust for observed effect measures. While adjustment procedures may contribute to understanding of the results of a study, caution should be exercised in interpreting the study results if such adjustment procedures have been applied (Armstrong et al., 1992). First, the assumptions underlying the adjustment formulae may not be appropriate. In particular, an assumption about non-differential exposure measurement error may not be correct. Also the estimates of measurement error should based on the same study population, although these error

estimates are generally not available so that estimates derived from other populations are applied. Finally, in many instances, no definitely accurate measure is available for judging validity. Information from reliability studies can be used in adjustment procedures, to the extent that the reliability study provides information about the validity of measured variables. However, reliability studies may yield only an upper limit for the validity coefficient. This leads to a conservative estimate of the true relative risk under non-differential measurement.

The formulae for adjusting for measurement errors can be very complex if an attempt is being made to adjust for differential and non-differential error relating to both exposure and health outcome measures. These formulae are beyond the scope of this textbook and, in fact, are rarely used because of their complexity. Instead, formulae are presented here for adjustment of non-differential exposure measurement error when it is assumed that disease outcome has been measured accurately. These formulae are essentially the same as those presented above, but the formulae are rearranged to produce the "true" effect measure as a function of the observed effect measure and the validity coefficient estimate derived from the validation study.

An estimate of the true effect measure can be made on the basis of the observed effect measure if the correlation between the observed and the "true" exposure (ρ_{TX}) is known (Armstrong at al., 1992). The validity correlation coefficient (ρ_{TX}) is estimated (r_{TX}) from a validation study in which the measure used for the study is compared to an accurate (and usually more costly) measurement technique in a sample of the study population. If a known accurate measurement technique is not available (i.e. validity), the correlation between two alternative measurement techniques (i.e. reliability, r_{x1x2}) may be used to estimate the bounds for an adjustment procedure.

Continuous exposure, continuous outcome (non-differential error)

$$\beta_T = \beta_O / r_{TX}^2$$

Continuous exposure, dichotomous outcome (non-differential error)

$$OR_T = OR_O^{1/r_{TX}^2}$$

where,

$_T$ = true regression coefficient; β_O = observed regression coefficient;
r_{TX} = correlation between measured exposure and true exposure.

If a reliability study is carried out using two parallel measures with uncorrelated errors (assuming that the second measure, X_2, is at least as reliable as the first measure, X_1, then the reliability coefficient ρ_{x1x2} can be used to estimate the upper and lower bound for the validity coefficient of X_1. That is,

$$\rho_{X_1X_2} < \rho_{TX_1} < \sqrt{\rho_{X_1X_2}}$$

For example, in a case–control study of the effect of radon exposure on lung cancer, the observed odds ratio was 2.0 per unit lifetime exposure to radon. a sub-study was performed using two different measures of radon exposure and found a 0.8 correlation between the two measures. Thus the validity coefficient would be between 0.8 if the second measure had the same level of measurement error as the first measure and $\sqrt{0.8}$ or 0.894 if the measurements were perfectly valid. Consequently, the bounds of the true odds ratio would be 2.38 to 2.95.

Dichotomous exposure, dichotomous outcome (differential or non-differential error)

$$OR_T = \frac{P_D(1 - P_N)}{P_N(1 - P_D)}$$

where,

$$P_D = (p_D - 1 + spec_D) / (sens_D + spec_D - 1)$$

and

$$P_N = (p_N - 1 + spec_N) / (sens_N + spec_N - 1)$$

Thus, the true odds ratio can be estimated if sensitivity and specificity of exposure classification are known for both disease and control groups.

The repeatability of categorical exposure variables can also be estimated from repeated measurements by calculating Cohen's kappa, a "chance-corrected" measure of agreement. For dichotomous exposure variables, it has been shown that the observed odds ratio in a study is related to the true odds ratio (Armstrong et al., 1992):

$$OR_T = \frac{(OR_O - 1)}{\kappa} + 1$$

For example, when the observed odds ratio is 2.5, and Cohen's kappa for the exposure variable is 0.75, a "true" odds ratio of 3.0 could be estimated.