Chapter 10: Use of epidemiological data

10.1 INTRODUCTION

Epidemiology and toxicology are complementary in risk assessment. Epidemiology is the direct human evidence component and, if based on sound epidemiological methods, can provide the most important evidence in characterising risk. Epidemiology is the principal driver in microbiological risk assessment, but it can assist the formal framework of EHRA at all of its stages.

The term 'risk' tends to be used in a subtly different and more specific way in epidemiology than in risk assessment. In epidemiology, risk describes the 'frequency of occurrence of a disease in one population compared with another, either as a difference in rates (attributable risk) or as a ratio of rates (relative risk)' (ACDP 1996 p. 20). The feature distinguishing the two populations by its presence or distribution is referred to as a 'risk factor'. The reliance on comparisons of disease rates between populations creates substantial limitations for the sensitivity of relative risk determination for common diseases (Thomas & Hrudey 1997).

Epidemiology is 'the study of the distribution and determinants of health-related states or events in specified populations, and the application of the study to the control of health problems' (Last 1988). A more recent description of epidemiology by the same author states:

> 'Epidemiology connects the dots, the isolated bits of information that begin to form a coherent pattern when connected in the right way ... The dots can come from anywhere. Identifying them demands a broader perspective, the ability to see the Big Picture ... Sometimes the way the dots are connected is instantly apparent. Sometimes painstaking investigation and analysis are required, for instance when the problem is a common but ill-defined condition ... caused by trace amounts of a highly reactive environmental toxin …' (Last 2010)

In simple terms, epidemiological methods compare health outcomes or calculated risk estimates in an exposed population or group, with those in a non-exposed population. The types of epidemiological studies that may be useful in making such evaluations are discussed briefly in Section 10.2.

Environmental epidemiology is considered to be a subspecialty of epidemiology that addresses the effects of environmental exposures on health and disease in the population (Baker & Nieuwenhuijsen 2008).

Epidemiological methods can be used to investigate the cause of adverse health effects, the natural history of health conditions and the description of the health status of populations, and to evaluate health-related interventions (Bonita et al. 2007). They also allow for the control for possible confounding factors that may influence the results. In the context of environmental health, epidemiological methods may also be used to characterise population exposures, investigate perceived clusters of disease, to develop health surveillance programs, to establish a baseline, and to monitor the consequences of risk management activities.

At the same time, there are often unrealistic expectations of what an epidemiological study may be able to achieve.

In epidemiological studies of the potential influence of environmental factors, the exposures encountered are rarely extreme, in contrast to the high levels of exposure that can be applied in animal-based toxicity tests. Such epidemiological studies are often confounded (see Section 10.3), and any associations found are commonly weak. Furthermore in large populations substantial effects in small numbers of sensitive individuals may be 'swamped' by a lack of effect in the majority.

Epidemiological studies are often most unreliable when measuring subtle effects, although these effects may be of public health significance when large populations are exposed. Even the most rigorously conducted studies are unreliable for detecting small increases in risk. Since it is impossible to prove an absence of risk from any human study, it is often considered that the principal value of epidemiology is to exclude major health effects of an exposure (NHMRC 2006).

The purpose of this chapter is to provide a basis for understanding the strengths and weaknesses of epidemiology in supporting risk assessment. As Mundt et al. (1998) noted that if the limitations of epidemiological studies are not understood by the risk assessment team, the validity of an assessment might be compromised by including inappropriate, possibly misleading, epidemiological data. In discussing the potential for epidemiological studies to throw up false negative outcomes, particularly in relation to associations with cancer, Boffetta et al. (2008) made a plea for epidemiologists to practise 'epistemological modesty' in the types of claims or conclusions drawn from their studies.

The systematic appraisal of epidemiological studies is intended to answer the question: 'Is there any other way of explaining the set of facts before us [i.e. the study results, is there any other answer equally, or more, likely than cause and effect' (WHO 2000a). Alternative explanations may result from chance, bias and confounding.

10.2 TYPES OF EPIDEMIOLOGICAL STUDY

Broadly speaking, epidemiological activity can be either 'descriptive' (reporting and describing the distribution of exposure and effect) or 'analytical' (designed to analyse and understand the degree of association between exposure and effect). Descriptive studies include case reports, case series and cross-sectional surveys. Cross-sectional surveys examine exposure and disease prevalence at the same point in time (or over a short duration of time) and thus are unable to support causal inference because it is not possible to know if exposure pre-dated the onset of disease. Similarly, case reports and cross-sectional surveys are unable to support causal inference.

In practical terms in environmental epidemiology, there are four main categories of analytical study (from Moolgavkar et al. 1999):

- case-control studies
- cross-sectional studies
- cohort (longitudinal) studies
- ecological studies (including a special sub-group known as time-series studies).

Cohort, cross-sectional and case-control studies differ from ecological studies in that information on exposure and disease is available on an individual basis. With ecological studies this information is only available on a group basis, so the community or region is the unit of analysis.

In **case-control studies**, exposure and other attributes of cases of the disease under investigation are compared with those from a suitable control or comparison group of persons unaffected by the disease, and analysed to yield effect estimates. The approach is to start with a diseased group and look backwards (retrospective cohort) to their history of exposures, and compare these exposures with exposures among a non-diseased cohort. The selection of appropriate controls to avoid bias is a significant challenge with case-control studies. However, among their advantages, case-control studies are relatively inexpensive, ideal for studying rare diseases and useful for investigating multiple, different exposures (Gregg 1996).

Cross-sectional studies measure the prevalence of disease and measure exposure and effect at the same time. They are relatively easy and economical to conduct and are particularly useful for measuring fixed characteristics of individuals such as socioeconomic status (Bonita et al. 2007).

Cohort studies follow cohorts or groups of individuals, defined in terms of their exposures, over time to see if there are differences in the development of new cases of the disease of interest (or other health outcomes) between the groups with and without exposure. Such studies can be carried out by either reviewing past records (retrospective) or by tracking people into the future (prospective cohort). The essential feature of these longitudinal studies is that for each individual prior exposure information can be related to subsequent disease experience (Breslow & Day 1987).

Ecological studies involve investigating a group of people such as those living within a geographical area (e.g. a region, state or territory). For example, place and time of residence can provide aggregate exposures, so may be used to create surrogate measures of the real exposure of interest (Elliott et al. 1992). Rates of disease and average exposure levels to a particular agent are determined independently, and on a group basis. This may give rise to spurious apparent correlation, called the 'ecological fallacy'. Because it is not ascertained whether individuals who have been exposed to the agent are the same individuals who developed the disease, statements about causal associations are inappropriate.

However, ecological studies are relatively inexpensive for linking available health datasets and environmental information and are useful for hypothesis-generation (Yassi et al. 2001). Examples of ecological studies are the assessments of the relationship between tobacco sales in different countries and lung cancer rates, and fluoride in water supplies and dental caries.

A subset of ecological studies, known as 'time series studies', is regarded as very helpful in understanding the influence of short-term fluctuations in air pollutants on day-to-day changes in population morbidity and mortality after controlling for factors such as season and air temperature. However, disentangling the effects of individual pollutants as measured in a mixture such as urban air pollution can be quite difficult.

To strengthen the design of ecological studies, Nurminen (1995) recommended selecting areas with populations that:

- are homogeneously exposed (to minimise within-area exposure variation)
- represent different extremes of exposure distribution (to maximise between-area exposure variations)
- are comparable with respect to co-variate distributions (e.g. socioeconomic status, demography)
- use the smallest possible sampling units for ecological analysis.

The largest number of environmental epidemiology studies found in the literature are of the ecological or cross-sectional type, because they are easier to carry out and cost less (Thomas & Hrudey 1997). However, as noted above and discussed further below in relation to assessing causality, such studies may be useful for identifying potential hazards or hypothesis generation, but they cannot determine cause and effect.

Characteristics of the various study types are summarised in Table 13.

Table 13: Study designs in environmental epidemiology that use the individual as the unit of analysis

Study design	Population	Exposure	Health effect	Confounders	Problems	Advantages
Case reports, case series and other descriptive studies	Community or various sub-populations	Records of past measurements	Mortality and morbidity statistics; case registers; other reports	Difficult to sort out	Hard to establish exposure-effect relationships	Cheap; useful to formulate hypotheses
Cross-sectional study	Communities or special groups; exposed versus non-exposed	Current	Current	Usually	Current exposure may be irrelevant to current disease	Can be done quickly; can use large populations; can estimate prevalence
Case-control study	Diseased (cases) versus non-diseased (controls)	Records or interview	Known at start of study	If confounders can be identified and measured they may be addressed	Difficult to generalise; may incorporate biases; cannot derive rates Recall bias is a major problem	Relatively cheap and quick; particularly useful for studying rare diseases
Time-series study	Large community (several million people); susceptible groups such as asthmatics	Current (e.g. daily) changes in exposure	Current (e.g. daily) variations in mortality	Often difficult to sort out; e.g. effects of influenza	Many confounding factors; often difficult to measure	Useful for studies on acute effects
Historical (retrospective) cohort study	Special groups; workers, patients, insured persons	Records of past measurement	Records of past or current diagnosis	Often difficult because of retrospective nature; depends on availability of previously obtained data	Need to rely on records that may not be accurate	Less expensive and quicker than a prospective study; can be used to study exposures that no longer exist
Prospective cohort study	Community or special groups; exposed versus non-exposed	Defined at outset of study (can change during study)	To be determined during study	Usually easy to measure	Expensive and time consuming; exposure categories can change; high dropout rate possible	Can estimate incidence and relative risk; can study many diseases in one study; can describe associations that suggest cause–effect relationships
Experimental (clinical/ intervention) study	Community or special groups	Controlled or already known	To be measured during study	Can be controlled by randomisation of subjects	Expensive; ethical considerations; study subjects compliance required	Well accepted results; strong evidence for causality or efficacy of intervention

Adapted from: WHO (1991).

Epidemiological studies are rarely definitive, and a single epidemiological study cannot establish causality. Any such study is necessarily a sample of the total population of interest, so there will always be questions about being able to generalise from an individual study sample to the total population. Moreover, undetected methodological biases can only be overcome by having numerous studies by different investigators with different population samples. If consistency of outcome is demonstrated across many studies, the causal hypothesis becomes more likely. A 'weight of evidence' (WoE) approach is generally required, involving the interpretation of all available information and consideration of the strengths and weaknesses of each study.

Unfortunately, experimental interventions such as randomised controlled trials are rarely available to assist environmental health risk assessment, as it is not ethical to purposely expose populations to hazardous risks. It is appropriate, however, to expose groups to a lesser level of risk via a clean-up intervention. An example of an experimental intervention is a randomised trial of lead abatement procedures undertaken in Broken Hill (S. Corbett, personal communication). Epidemiological studies, depending on their design, may serve two purposes: hypothesis generation or assessment of a causal relationship. Their ability to evaluate a causal relationship may be limited by a lack of control of potential confounders or a lack of power (usually the result of limited sample sizes) (Samet et al. 1998).

10.2.1 Observational studies

Different observational study designs have different applications, advantages and disadvantages (see Tables 14 and 15). These comparisons assume the different types of studies are equally well designed. Even so, design variations may affect their performance and provide exceptions. Bonita et al. (2007) provides a more detailed description.

Table 14: Applications of different observational study designs

	Ecological	Cross-sectional	Case-control	Cohort
Investigation of rare disease	++++	–	+++++	–
Investigation of rare cause	++	–	–	+++++
Testing multiple effects of cause	+	++	–	+++++
Study of multiple exposures and determinants	++	++	++++	+++
Measurement of time relationships	++	–	+[a]	+++++
Direct measurement of incidence	–	–	+[b]	+++++
Investigation of long latent periods	–	–	+++	+[c]/-

Key:

+ to +++++ indicates the degree of suitability from least to most suitable; – indicates not suitable

a if prospective; b if population-based; c if retrospective or historical cohort study.

Table 15: Advantages and disadvantages of different observational study designs

	Ecological	Cross-sectional	Case-control	Cohort
Probability of:				
Selection bias	N/A	Medium	High	Low
Recall bias	N/A	High	High	Low
Loss to follow-up	N/A	N/A	Low	High
Confounding	High	Medium	Medium	Low
Time required	Low	Medium	Medium	High
Cost	Low	Medium	Medium	High

Tables 14 and 15 adapted from: Bonita et al. 2007.

10.3 INVESTIGATION OF APPARENT CLUSTERS

The assessment of an apparent cluster of non-communicable disease is a complex and resource-intensive task. It commonly involves investigation of a number of reported cases of cancer, or some other adverse health effect likely to be linked to an environmental exposure. It should be managed using a multidisciplinary approach using standardised analytical tools. The trigger for a cluster investigation is often the 'perception' that the incidence of the disease is unusually high in a region or scenario linked to a possible environmental exposure source (e.g. near a waste dump, or in an occupational setting). The assessment is usually first centred on whether the observed number of cases is consistent with that expected from the background incidence, or whether there is a sufficiently common pattern to the nature of the cancers or other health effect.

Specific guidance on cluster investigation has been developed by Queensland Health (2009), including suggested criteria for decision making. The NHMRC is working towards the development of guidelines that can be adopted nationally.