Epidemiology

Epidemiology is the study and analysis of the patterns, causes, and effects of health and disease conditions in defined populations. It is the cornerstone of public health, and shapes policy decisions and evidence-based practice by identifying risk factors for disease and targets for preventive healthcare. Epidemiologists help with study design, collection, and statistical analysis of data, amend interpretation and dissemination of results (including peer review and occasional systematic review).

Epidemiology has helped develop methodology used in clinical research, public health studies, and, to a lesser extent, basic research in the biological sciences.

Major areas of epidemiological study include disease etiology, transmission, outbreak investigation, disease surveillance, forensic epidemiology and screening, biomonitoring, and comparisons of treatment effects such as in clinical trials. Epidemiologists rely on other scientific disciplines like biology to better understand disease processes, statistics to make efficient use of the data and draw appropriate conclusions, social sciences to better understand proximate and distal causes, and engineering for exposure assessment.

Case–control studies –

A case–control study involves the identification of individuals with ('cases') and without ('controls') a particular disease or condition. The prevalence (or level) of exposure to a factor is then measured in each group. If the prevalence of exposure among cases and controls is different, it is possible to infer that the exposure may be associated with an increased or decreased occurrence of the outcome of interest (see Section 9.5).

Example 9.1. *The relationship between use of conjugated estrogens and the risk of endometrial cancer was examined among 188 white women aged 40–80 years with newly diagnosed endometrial cancer and 428 controls of similar age hospitalized for non-malignant conditions requiring surgery at the Boston Hospital for Women Parkway Division, Massachusetts, between January 1970 and June 1975. The data on drug use and reproductive variables were extracted from hospital charts and from the medical records of each woman's private physician. Thirty-nine per cent of the cases and 20% of the controls had used conjugated estrogens in the past (Buring et al., 1986).*

9.1 Study hypothesis

As with any other type of study, the specific hypothesis under investigation must be clearly stated before a case–control study is designed in detail. Failure to do this can lead to poor design and problems in interpretation of results. Case–control studies allow the evaluation of a wide range of exposures that might relate to a specific disease (as well as possible interactions between them). clearly illustrates this feature.

Example 9.2. *A population-based case–control study was carried out in Spain and Colombia to assess the relationship between cervical cancer and exposure to human papillomavirus (HPV), selected aspects of sexual and reproductive behaviour, use of oral contraceptives, screening practices, smoking, and possible interactions between them. The study included 436 incident*

cases of histologically confirmed invasive squamous-cell carcinoma of the cervix and 387 controls of similar age randomly selected from the general population that generated the cases (Muñoz et al., 1992a).

9.2 Definition and selection of cases:-

9.2.1 Case definition

Precise criteria for the definition of a case are essential. It is usually advisable to require objective evidence that the cases really suffer from

the disease or condition of interest, even if, as a result, some true cases have to be eliminated. For instance, a histologically confirmed diagnosis should be required for most cancers. By accepting less well documented cases, the investigator runs the risk of diluting the case group with some non-cases and lessening the chances of finding real exposure differences between cases and controls. It is sometimes impossible to eliminate all cases whose diagnosis is not properly documented, particularly if the pool of available cases is relatively small. In these circumstances, it may be possible to classify the cases according to diagnostic certainty. Such classification allows assessment of the extent to which the results are likely to be affected by disease misclassification (see Chapter 13). Suppose, for instance, that cases in a particular case–control study are classified as 'definite', 'probable' or 'possible'. If there is disease misclassification, a gradual decline in relative risk from the 'definite' to the 'possible' category should become apparent in the analysis, since the probability that non-cases may have been misdiagnosed as cases increases from the 'definite' to the 'possible' category. The case definition should be established in such a way that there is no ambiguity about types of cases and stages of disease to be included in, or excluded from, the study. The choice of cases should be guided more by concern for validity than for generalizability. For example, in a study of breast cancer, we may learn more by limiting the cases (and the controls) to either pre- or post-menopausal women than by including women of all ages (unless the number of cases in each group is large enough to allow separate analyses), since the risk factors for pre- and post-menopausal breast cancers may be different. By ensuring that the cases are a relatively homogeneous group, we maximize the chances of detecting important etiological relationships. The ability to generalize results to an entire population is usually less important than establishing an etiological relationship, even if only for a small subgroup of the population. Cases should also be restricted to those who have some reasonable possibility of having had their disease induced by the exposure under investigation.

Example 9.4. *A multinational, hospital-based case–control study was conducted to evaluate the relationship of combined oral contraceptive use to the risk of developing five different site-specific cancers. The study was conducted in 10 participating centres in eight countries (Chile, China, Colombia, Israel, Kenya, Nigeria, Philippines and Thailand) from October 1979 to September 1986. Women with newly diagnosed cancers of the breast, corpus uteri, cervix uteri, ovary and liver were eligible if born after 1924 or 1929 (depending on when oral contraceptives became locally available) and had been living in the area served by the participating hospital for at least one year (WHO Collaborative Study of Neoplasia and Steroid Contraceptives, 1989).*

9.3 Definition and selection of controls:-

9.3.1 Definition of controls

Controls must fulfil all the eligibility criteria defined for the cases apart from those relating to diagnosis of the disease. For example, if the cases are

women with breast cancer aged 45 years and over, the controls must be selected from women in the same age group without the disease. If the disease being studied is uncommon in the group serving as a source of controls, little, if any, diagnostic effort or documentation is needed to rule out the disease in the selected controls. A simple interview question will often suffice. However, if the disease is common, a greater effort to minimize misclassification, such as a review of the individuals medical records, is desirable (as in example 9.9).

Example 9.9. *In the cervical cancer case–control study mentioned in Example 9.2, controls were eligible if they were 70 years of age or younger, had not received previous treatment for cervical cancer or had not been hysterectomized, and if the cytological smear taken at the time of recruitment was normal or had only inflammatory changes (Pap classes I and II) (Muñoz et al., 1992a)*

9.4 Measuring exposures

Data on the relevant exposures can be obtained by personal, postal or telephone interview, by examining medical, occupational or other records, or by taking biological samples. Whatever method is chosen, it is fundamental to ensure that the information gathered is unbiased, i.e., it is not influenced by the fact that an individual is a case or a control. Ideally, the investigator or interviewer should be 'blind' to the hypothesis under study and to the case/control status of the study subjects. In practice, this may be difficult to accomplish, but all possible efforts should be made to ensure unbiased collection of data to minimize observer bias. Particular effort is required in multicentric studies to ensure standardization of data collection techniques across the different participating centres. Bias can also occur when the validity of the exposure information supplied by the subjects differs for cases and controls (responder bias). Subjects with a serious disease are likely to have been thinking hard about possible causes of their condition and so cases may be inclined to give answers that fit with what they believe (or think is acceptable to say) is the cause of their illness. This type of responder bias is called recall bias. Responder bias can be minimized by keeping the study members unaware of the hypotheses under study and, where possible, ensuring that both cases and controls have similar incentives to remember past events.

9.5 Analysis

The analysis of data from case–control studies depends on their design. Individual-matched studies require a different type of analysis from ummatched (or frequency-matched) studies.

9.5.1 Unmatched (and frequency-matched) studies

The first step in the analysis of an unmatched case–control study is to construct a table showing the frequency of the variables of interest separately for cases and controls. The frequency of some of these variables in the controls may help to judge whether they are likely to represent the

population from which the cases arise. For instance, in Example 9.15 , the distribution of schooling, parity, smoking, etc. in the control group of this population-based study may be compared with governmental statistics or results from surveys conducted in the same areas

In Example 9.15, the distributions of some of the variables known to be risk factors for cervical cancer are consistent with those found in other studies in that cases were more likely to have a lower educational level, higher parity and a greater number of sexual partners than controls. They were also more likely to have ever used oral contraceptives or smoked.

9.5.2 Individual-matched studies

Individual-matched studies require a special type of analysis, in which the 2 × 2 table takes a different form. Let us consider the simplest situation where there is only one control per case. The status of the cases with regard to the presence or absence of the exposure of interest is cross-tabulated against the exposure status of their respective controls (Table9.8)

Table 9.8.
Layout of a 2 × 2 table with data from an individual-matched case–control study (control-to-case ratio = 1:1).

		Controls		Total
		Exposed	Unexposed	
Cases	Exposed	*r*	*s*	*a*
	Unexposed	*t*	*u*	*b*
	Total	*c*	*d*	*N*/2

In this table, *r*, *s*, *t*, *u* represent the number of pairs in which
r = case exposed and control exposed (+ +)
s = case exposed but control not exposed (+ –)
t = case not exposed and control exposed (– +)
u = case not exposed and control not exposed (– –)
The marginal totals (*a*, *b*, *c*, *d*) of this table correspond to the entries in the cells of the table for the unmatched studies. The total for the entire table is *N*/2 pairs, where *N* represents the total number of paired individuals.
The *matched odds ratio* can be calculated as
Odds ratio = *s*/*t* (provided *t* is not equal to 0)
This odds ratio calculation considers only the discordant pairs. It can be explained intuitively: pairs where both case and control were exposed or where both were unexposed give no information about the relationship of

the exposure to disease (Example 9.18)

The analysis is more complex than shown here if there is more than one

control per case (see Breslow & Day (1980), chapter 5).

9.6 Interpretation of results

Case–control studies are well suited to study diseases of long induction, because no lengthy follow-up is involved. They are also suitable for studying rare diseases, since a prospective

cohort study would require the recruitment of a very large number of individuals and a long follow-up period to ensure the accrual of a sufficient number of cases.

The interpretation of case–control studies is, however, less straightforward than that of cohort studies and the investigator must always consider.

Example 9.18.

A case–control study was carried out in Canada to assess whether artificial sweeteners, particularly saccharin, increased the risk of bladder cancer. Newly diagnosed cases of bladder cancer that occurred
among residents in the provinces of British Columbia, Nova Scotia and Newfoundland between April 1974 and June 1976 were identified through provincial cancer registries and cooperative pathologists and urologists. A total of 821 eligible cases were ascertained, and 632 of these were personally
interviewed in their homes using a structured questionnaire. Reasons for failure to interview included death (56), refusal (65), too ill to be interviewed (25), and refusal of permission by the attending physician (34). Most interviews were done within three months of diagnosis, and all within six
months. For each case, an individual matched on sex, age (within 5 years), and neighbourhood residence was interviewed (Howe et al*., 1977). The main results are shown in Table 9.9.*