FOUNDATIONS OF EPIDEMIOLOGY

Abraham M. Lilienfeld, M.D., M.P.H., D.Sc.
Department of Epidemiology
The Johns Hopkins University
School of Hygiene and Public Health

New York
Oxford University Press
1976

8 Retrospective and Cross-sectional Studies

A. GENERAL DESCRIPTION OF METHOD

After a statistical association between a characteristic and a disease has been observed by mortality and/or morbidity studies in a community or population group, the epidemiologist attempts to confirm it by determining whether this association is also present within individuals with the disease as compared to those without the disease. Clinical and/or experimental observations often suggest an association which also requires confirmation by such means. Retrospective or cross-sectional studies are usually used to determine whether such associations are present within individuals. This chapter deals with both types of studies, which are frequently used together in actual practice because of their methodological similarities. There are, however, differences in the inferences that may be derived from them which will be discussed.

In cross-sectional and retrospective studies, comparisons are made between a group of persons who have the disease and a group which does not; those with the disease are called "cases" and those without are generally called "controls." This is the reason for the usual description of these studies in the literature as "case-control" studies. The control group has also been called a "comparison group," to distinguish it from the control group used in experiments. Cross-sectional studies have also been called "prevalence" studies, and retrospective studies, "backward-looking," or "case history" studies.

The major distinction between cross-sectional and retrospective

studies is that, in the former, the characteristics being compared are present in the cases and controls at the time of the study, while in the latter, an attempt is made to determine whether the characteristics were present in the individuals in the past; usually, this is done by interview or a review of records. In both instances, the proportion of cases exposed to, or possessing, the characteristic (or factor) of etiologic interest is compared to the corresponding proportion in the control group. If a higher frequency of individuals with the characteristic is found among the cases than the controls, an association between the disease and the characteristic is indicated.

The following is an illustration of the difference in inferences that can be derived from cross-sectional and retrospective studies. One is interested in the relationship between myocardial infarction and the level of serum cholesterol. A group of coronary heart disease cases and a control group are selected, and blood is drawn to determine the serum cholesterol level. If the serum cholesterol level is higher among the cases than the controls, one can then state that an association exists between myocardial infarction and an elevated serum cholesterol level. In this cross-sectional approach, however, one does not know whether the elevated serum cholesterol preceded the onset of the myocardial infarction or followed it. If the latter were true, the elevated serum cholesterol could not be regarded as being etiologically important in the development of myocardial infarction. In contrast, in a retrospective study, one would seek information on the serum cholesterol level before the onset of the myocardial infarction; if an association is established, a causal inference is much stronger.

This distinction between cross-sectional and retrospective studies mainly applies when the selected study factor is a physiological or biochemical characteristic of the individual since these characteristics are more likely to have been influenced by the disease itself. It is, however, possible that the disease may have been responsible for the person's change in environment after the disease had occurred. Therefore, the distinction between cross-sectional and retrospective studies must be kept in mind even when environmental factors are studied. When interested in determining whether prior exposure to an environmental factor is etiologically important, the investigator will attempt to obtain a history of such exposure by interviewing the cases and controls, although, in practice, information on both current and past characteristics is usually gathered. One must

Table 8–1. Framework of a Retrospective Study

Characteristic	Number of Individuals		
	With Disease (Cases)	Without Disease (Controls)	Total
With	a	b	a + b
Without	c	d	c + d
Total	a + c	b + d	a+b+c+d = N

constantly be aware that the derivation of inferences depends upon the temporal sequence between the characteristic and disease. In contrast to cross-sectional studies, retrospective studies should be regarded as "time-span" studies.

The basic data from a retrospective study is shown in Table 8–1. If $\frac{a}{a+c}$ is statistically significantly greater than $\frac{b}{b+d}$, an association can be said to exist between the disease and the characteristic. Table 8–2 provides an example, using data from Table 1–2 on the number of cigarette smokers among bladder cancer patients and a control group. The percentage of cigarette smokers among the bladder cancer patients is 60 per cent $(\frac{192}{321})$ compared to 46 per cent $(\frac{156}{337})$ among the controls. The methods for determining the statistical significance of these different percentages can be found in

Table 8–2. Number of Cigarette Smokers and Non-cigarette Smokers among Bladder Cancer and Control Patients Admitted to Roswell Park Memorial Institute, 1945–1955

Smoking Habit	Number of Individuals		
	With Bladder Cancer	Controls*	Total
Cigarette smokers	192	156	348
Non-cigarette smokers	129	181	310
Total	321	337	658

* No disease group in Table 1–2.

many biostatistic textbooks (1, 12, 35). Many of the examples in this chapter will be concerned with cigarette smoking and lung cancer since several of the study methods were developed to investigate this relationship.