Chapter 10

# Clinical Studies of Disease Outcome

Just as some questions relating to disease occurrence and disease etiology are best answered by studying population groups, clinical problems often require the study of *groups* of patients. Many methods for studying patient groups are similar to the epidemiologic methods for studying populations, discussed in previous chapters.

The process by which healthy people become sick and the factors that determine who will become sick and who will stay healthy are the primary concern of epidemiology. Many clinical studies, on the other hand, aim at sick people and try to identify the factors that determine what the outcome of illness will be. This difference in focus between the two types of studies is illustrated in Fig. 10-1. Note that illness or disease can have several outcomes, including recovery, improvement, no change, worsening, complications, disability, and death.

The ultimate goal of epidemiology is to learn how to prevent





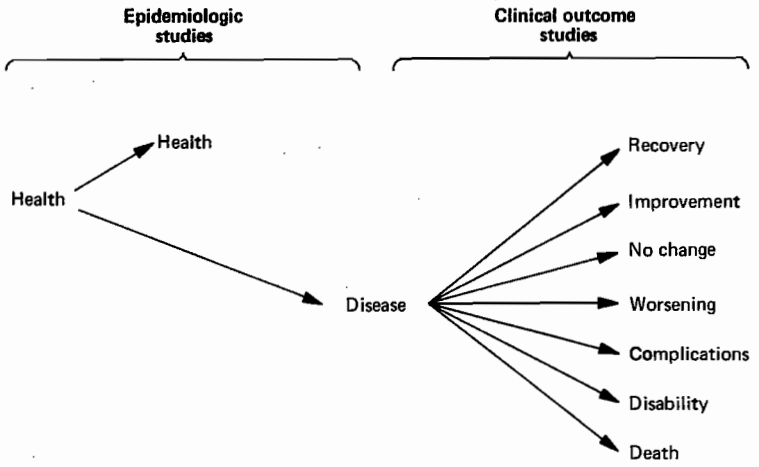



**Figure 10-1** Areas of concern of epidemiologic studies and clinical outcome studies.

disease. The ultimate goal of clinical studies is to learn how to cure or successfully treat disease.

The purpose of this chapter is to demonstrate some of the parallels between clinical studies of disease outcome and epidemiologic studies and to describe the analytic methods commonly used to measure disease outcome.

## Natural History of Disease

Studies of the *natural history of disease* are analogous to descriptive studies in epidemiology. The outcomes of a particular disease are observed and the proportions of the affected patients developing each outcome are measured. This information is the basis of *prognosis*, that is, predicting a patient's future. As in descriptive epidemiologic studies, disease outcomes are generally determined for major subgroups of patients such as males versus females, various age groups, and so on.

A good example of a study of the natural history of disease is Bland and Jones' (1951) 20-year study of 1,000 children and adoles-

cents with rheumatic fever or chorea. These patients, initially hospitalized at the House of the Good Samaritan in Boston, were carefully followed up into adulthood. Among the findings were that 65 percent of the children had signs of rheumatic heart disease when they recovered from their acute illness, but 16 percent of those with such signs had no evidence of heart disease 20 years later. On the other hand, 44 percent of those without apparent heart disease initially had valvular disease when they were examined as adults. Also described were the recurrence rates of acute rheumatic fever, the evolution of murmurs, and the frequency of deaths and other sequelae of the disease.

### Analytic Studies

The clinical investigator usually wishes to go beyond general descriptions of prognosis and to determine what factors lead to improvement, worsening, death, and other outcomes. Such factors include patient characteristics and environmental influences. One of the main environmental factors that is investigated is, of course, therapy.

Analytic clinical investigations of prognostic factors may be carried out in a fashion quite analogous to prevalence, case-control, and incidence studies in epidemiology. A physician is conducting what amounts to an informal prevalence study when he makes rounds on two wards caring for paralyzed stroke patients and notices that in one ward, several patients have decubitus ulcers (bedsores) and on the other, the patients are ulcer-free. He will probably conclude that being on the first ward is conducive to the development of this complication of paralytic stroke and will make some appropriate comments to the nursing staff.

Analytic studies of factors affecting prognosis are usually similar to incidence studies. That is, attributes of the group of patients are assessed early in the course of the illness. Then, the patients are followed up to determine outcome.

The clinical investigator can adopt this prospective follow-up approach much more readily than can the epidemiologist. The rates of development of many disease outcomes are relatively high, compared to the incidence of most diseases in a population. Thus, a



relatively small patient group can be observed in one clinic or hospital until the various outcomes are noted.

Consider, for example, the follow-up study by Stahlman et al. (1967) to determine characteristics predicting the outcome of hyaline membrane disease in the newborn. Of 115 affected newborns studied, 33, or 29 percent, died in the neonatal period. A number of measurements taken within 12 hours of birth, such as arterial-blood oxygen tension, birth weight, and respiratory rate, all proved to be related to mortality, and statistical-significance tests showed that these relationships could not reasonably be attributed to chance. Thus, the predictive value of these measurements was demonstrable in this study of only several dozen patients.

Some analytic follow-up studies of prognosis deal with events that develop relatively slowly and infrequently, so that large numbers of patients must be followed for years. This is particularly true of chronic diseases. The Health Insurance Plan of Greater New York (HIP) has been investigating the prognosis of patients with angina pectoris and myocardial infarction. One such study demonstrated a relationship of blood pressure in these patients to the probability of subsequent myocardial infarction and cardiac death—the higher the blood pressure, the worse the prognosis. This study was based on 275 cases of angina pectoris and 881 cases of a first myocardial infarction found among 55,000 men during a 4-year case-finding period. The cases were followed up for 4.5 years (Frank et al., 1972).

When an analytic follow-up study cannot be carried out, it may be practical to use an approach analogous to the case-control method in epidemiology. That is, a group of patients with one particular outcome may be compared with a group showing another outcome, to see whether the two groups differ in any characteristic that might have affected or predicted the outcome. An example is Ellenberg's (1971) study of sexual impotence complicating diabetes mellitus. Forty-five impotent diabetic men ("cases") were compared with thirty male diabetics who were not impotent ("controls"). The potent diabetics were selected to match the impotent group with respect to age distribution and duration of diabetes. The striking difference between the two groups was in the percentage showing evidence of neuropathy affecting the autonomic system—82 percent of the impotent versus 10 percent of the potent. Thus it could be

concluded that most cases of impotence in diabetics were due to diabetic neuropathy rather than endocrine or other abnormalities.

### Therapeutic Trials

The therapeutic trial is an experiment as applied to clinical medicine. In it, a drug, a surgical operation, or other therapy is applied to patients and the outcome is compared with that observed in a suitable control group.

It is essential that alternative therapies be evaluated in a well-controlled fashion using, whenever possible, the techniques of random allocation and blind assignment and assessment described in Chap. 9. The influence of the therapist's personality and the placebo effect (or tendency of patients to respond favorably even when a drug has no active ingredients) are potent determinants of outcome and should not be allowed to bias the experiment. Furthermore, because of wide variations in the way individual patients respond to treatment, large groups of patients are often required. Large groups will help ensure that an observed relationship between a treatment and an outcome is not due to chance and that the relationship has some general applicability.

The value of large patient series is apt to be forgotten by clinicians working with patients on an individual basis. A physician's use or avoidance of a particular therapy is often guided by his experience with a few patients. His view of the values or dangers of a particular treatment may be exaggerated just because, as luck would have it, the first two or three patients treated happened to respond unusually well or unusually poorly.

There is a widespread belief that the individual physician is the best judge of the value of a drug or other treatment. Through his knowledge of the patient, he may well be the best judge of what is most appropriate for that patient's particular problems. However, the average physician's limited experience with a few patients does not usually provide enough information to state a general principle or conclusion that one therapy is better than another. He may be able to detect dramatic effects such as the value of penicillin versus no antibiotic in treating lobar pneumococcal pneumonia. But conclusions as to less-striking differences between therapies should be



based on good-sized and representative series of patients with observations controlled as well as possible.

Medical history is full of examples of therapies which become accepted or popular in an epidemic of enthusiasm based on uncontrolled observations. Feeding this epidemic is the preference of authors and journals for reporting positive findings over negative findings. If the treatment is either not helpful or actually harmful, its use may eventually diminish or end, as its deficiencies become recognized. Unfortunately, during the period of general acceptance, withholding the treatment from some individuals, as is required in a well-controlled experiment, may be considered unethical. Thus it is important to perform a good therapeutic trial as early as possible after the therapy is developed.

Nevertheless, controlled trials are better carried out late than never. For example, the Boston Inter-Hospital Liver Group (BILG) recently completed a well-controlled therapeutic trial which failed to confirm the long-term value of a widely accepted surgical treatment (Resnick et al., 1969). Portacaval-shunt operations had been carried out as an elective prophylactic measure on patients with cirrhosis of the liver to relieve the excess pressure in esophageal varices and prevent serious bleeding episodes. Acceptance of the procedure by the medical profession was based on uncontrolled observations that cirrhotic patients who received this operation did better and lived longer than those who did not. What is often forgotten is that surgeons naturally prefer to operate on the relatively healthy or good-risk patients and reject the poor-risk patients as operative candidates.

In the BILG study, 93 cirrhotic patients with esophageal varices and no prior major bleeding episodes were randomly divided into a surgical and medical group. To avoid selection of the better-risk candidates for shunt surgery in this experiment, each patient was randomly assigned *after* the physicians and surgeons agreed that he or she was a candidate and *after* the patient had consented to have surgery. Both groups were followed up for several years.

The operation apparently did prevent bleeding episodes, as there were significantly more patients with subsequent hemorrhages in the medical group (12/45) than in the surgical group (1/48). However, the mortality of the surgical and medical patients was quite

similar. Although the surgical patients were less apt to die of bleeding, they were more apt to die of the hepatorenal syndrome. They were also more prone to develop hepatic encephalopathy.

Another recent controlled therapeutic trial did confirm the value of a much-used but still-debated treatment. For many years, even the individual practitioner could reliably observe that antihypertensive drug therapy brought about a dramatic improvement in the prognosis of severe and malignant hypertension. However the value of drugs for mild to moderate hypertension was less easy to recognize and, until quite recently, was subject to considerable debate. As a result, the Veterans Administration (1967, 1970) carried out a cooperative study in which 523 men with diastolic blood pressures of 90 to 129 mm Hg were assigned randomly to active drug therapy or placebo. Before random assignment there was a trial period during which the potentially uncooperative subjects—those who did not attend clinic regularly or take at least 90 percent of a marked placebo—could be eliminated. (Because most hypertensives feel well, there is little immediate gratification for them in following a regular therapeutic program.)

Therapeutic benefit to the drug-treated group was apparent after only 20 months of follow-up of those starting with diastolic levels of 115 to 129 mm Hg. Only 1 of 73 treated patients developed a major cardiovascular-renal complication, as compared to 27 of 70 control subjects, of whom 4 died. One other treated patient exhibited drug toxicity and had to be removed from the study therapy.

Longer follow-up of more subjects was required to demonstrate benefits of treating milder hypertension—90 to 114 mm Hg diastolic pressure. A total of 380 patients were followed up for an average of 3.3 years. Major complications were observed in 56 of 194 controls, as compared to only 22 of 186 treated subjects. Some complications, such as stroke, showed a markedly lower incidence among the treated group.

Concomitant with the reporting of controlled observations such as these has been a growing awareness that hypertension is serious, and that large numbers of persons in this country are hypertensive and not aware of it. Moreover, many persons who are aware of hypertension are not being treated adequately or consistently. Thus



the detection and sustained treatment of hypertension may become a major public health effort in the near future.

## Commonly Used Measures of Disease Outcome

**Rates** Just as incidence rates are used in epidemiology to measure the development of disease in healthy persons, the outcomes of illness can be measured similarly in groups of sick persons. Thus one may speak of recovery rates, disability rates, death rates, and so on, referring to the proportion of the ill that recover, become disabled, or die per unit of time. Again, the proportion of the sick who manifest a particular outcome at one point in time is analogous to a point prevalence rate of disease in a general population.

**Survival** Measures of mortality outcome are often expressed in terms of *survival* rather than death. For comparative purposes, it is not particularly important whether one focuses on successes or failures. However, the data from clinical studies are so often analyzed and presented in terms of survival that it is desirable to be familiar with the approaches used. It should be remembered, also, that these measures need not be restricted to life and death. They can be applied to any mutually exclusive alternatives. Thus, in a study of the development of congestive heart failure in cardiac patients, remaining free of failure can be considered analogous to survival.

One of the most common measures of outcome is the proportion surviving for a particular duration. Any duration may be chosen—5 years is frequently used in studies of the surgical treatment of cancer, because for many types of cancer, if a patient survives for 5 years it is likely that he has been cured. Thus the "5-year survival rate" or "5-year cure rate" merely refers to that proportion of the original patient group still alive after 5 years of follow-up.

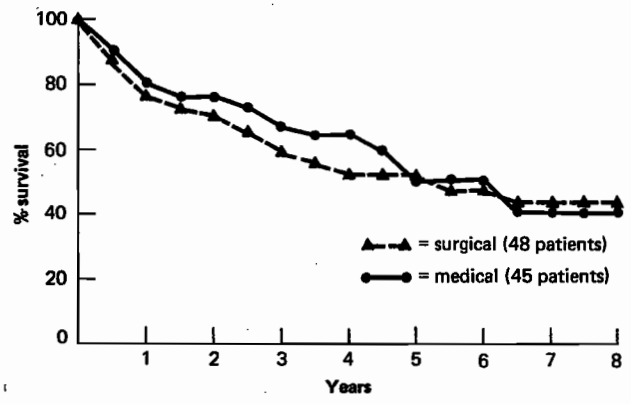
Another measure of survival that has been used is the "mean duration of survival." As mentioned in Chap. 2, page 19, the mean duration should be used for comparative purposes only when all

patients have died. When some are still living, it is preferable to compare *median duration of survival* or some other *quantile* of *survival durations* because once the stated percentage have died, their survival cannot change. For example, after 75 out of 100 patients have died, the survival duration of the seventy-fifth person becomes the 75th percentile of survival durations for the entire group. This cannot change no matter how much longer the other 25 live. The mean, on the other hand, is not finally determined until all 100 have died.

One of the most common and probably the most informative measures of survival is the survival curve. Starting initially at 100 percent, it shows the proportion still surviving at each subsequent point in time for as long as information is available. Fig. 10-2 shows the curves for the medical and surgical patients in the BILG study of portacaval shunt. The similarity in their survival experience is apparent.

Another graph, Fig. 10-3, shows marked differences in survival for several subgroups of patients with scleroderma, from the study by Medsger et al. (1971). The proportions of scleroderma patients surviving at the end of each year after entry into the study are shown by solid black circles. Those who had no involvement of their lungs,

**Figure 10-2** Survival of surgical and medical patients in the Boston Inter-Hospital Liver Group's controlled therapeutic trial of portacaval shunt surgery for esophageal varices. *(Reproduced, by permission, from Resnick et al., 1969.)*





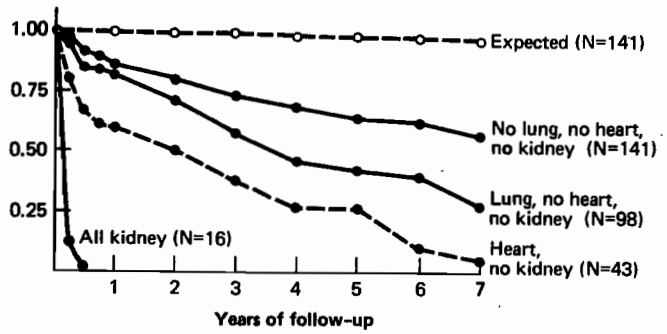




**Figure 10-3** Survival of scleroderma patients according to organ involvement. Ordinate shows proportion surviving. *(Reproduced, by permission, from Medsger et al., 1971.)*

heart, or kidneys did the best, with 56 percent still alive after 7 years. Subgroups with poorer survival were next, those with lung involvement; then, those with heart involvement; and finally, those with kidney involvement, all of whom died within the first half year. For comparison, the expected survival curve is shown on top with clear circles. This is the survival that would have been expected for a group of this age, sex, and racial composition if the overall United States mortality rates for the study years had been applicable.

Construction of survival curves for a certain duration following a specific event or time does not require that all patients be observed for that entire duration. Consider an example in which persons are to be followed for 10 years starting at the time their disease was first diagnosed. The experience of a person who moves away and is lost to follow-up after 5 years is still useful in determining survival rates for the first 5 years. Similarly, someone who is diagnosed and enters the study 1 year before the date that follow-up observations are to be completed contributes to those persons observed during the first year after diagnosis.

Thus, all persons who are observed during each unit of time measured from the starting event can contribute their experience to the survival-rate computation for that time unit. The so-called *actuarial* or *life-table* method takes advantage of all these observations by computing survival rates for each time unit and combining these rates together into one composite survival curve. For details as

to methods, which are not difficult to carry out, see Berkson and Gage (1950), Cutler and Ederer (1958), or Hill (1971).

**Importance of Starting Times** When survival curves (or mortality rates) of two groups are to be compared, it is important that both have the same starting point. The starting time may be placed at the onset of symptoms, the first diagnosis, the beginning of therapy, discharge from a hospital, or some other landmark in the course of the disease.

Failure to follow this principle has led to many conflicting claims and erroneous conclusions as to benefits of therapy. For example, two equally good surgical treatments will appear to have different results if survival is measured from the hospital discharge date for one, and from the date of operation for the other. Measuring from date of discharge excludes operative and immediate postoperative mortality.

Although the inclusion or exclusion of operative mortality makes for an obvious error, more subtle and hard-to-recognize biases may result when follow-up of two groups does not begin at strictly comparable times. Consider a study to evaluate the efficacy of a new procedure for the early diagnosis of a disease. Even if detecting the disease early does not prolong life, it might appear to do so if survival is measured from the date of *early* diagnosis instead of from the usual diagnosis date resulting from traditional methods. Procedures for overcoming this bias are discussed by Feinleib and Zelen (1969).

Similarly, treatment measures for rapidly fatal diseases may appear more effective than they really are if they are initiated after a short delay. Part of the apparent improvement in in-hospital mortality from myocardial infarction, experienced by patients in coronary-care units, may be related to the fact that many heart attack victims die shortly after the onset of the attack. As noted by Kodlin, patients in coronary-care units have already survived the short delay between admission to the hospital and admission to the unit.

## REFERENCES

Berkson, J., and R. P. Gage. 1950. Calculation of survival rates for cancer. *Proc. Staff Meet. Mayo Clinic*, **25**:270–286.



Bland, E. F., and T. D. Jones. 1951. Rheumatic fever and rheumatic heart disease: A twenty-year report on 1,000 patients followed since childhood. *Circulation*, **4**:836–843.

Cutler, S. J., and F. Ederer. 1958. Maximum utilization of the life table method in analyzing survival. *J. Chron. Dis.* **8**:699–712.

Ellenberg, M. 1971. Impotence in diabetes: The neurologic factor. *Ann. Intern. Med.*, **75**:213–219.

Feinleib, M., and M. Zelen. 1969. Some pitfalls in the evaluation of screening programs. *Arch. Environ. Health*, **19**:412–415.

Frank, C. W., E. Weinblatt, S. Shapiro, and R. Sager. 1972. Prognosis of men with coronary heart disease as related to blood pressure. *Circulation*, **38**:432–438.

Hill, A. B.: *Principles of Medical Statistics*, 9th ed. (London: Oxford University Press, 1971), pp. 228–236.

Kodlin, D. On the status of coronary care unit statistics. To be submitted for publication.

Medsger, T. A., A. T. Masi, G. P. Rodnan, T. G. Benedek, and H. Robinson. 1971. Survival with systemic sclerosis (Scleroderma): A life-table analysis of clinical and demographic factors in 309 patients. *Ann. Intern. Med.*, **75**:369–376.

Resnick, R. H., T. C. Chalmers, A. M. Ishihara, A. J. Garceau, A. D. Callow, E. M. Schimmel, E. T. O'Hara, and the Boston Inter-Hospital Liver Group. 1969. A controlled study of the prophylactic portacaval shunt: A final report. *Ann. Intern. Med.*, **70**:657–688.

Stahlman, M. T., E. J. Battersby, F. M. Shepard, and W. J. Blankenship: 1967. Prognosis in hyaline-membrane disease: use of a linear-discriminant. *New Engl. J. Med.*, **276**:303–309.

Veterans Administration Cooperative Study Group on Antihypertensive Agents. 1967. Effects of treatment on morbidity and mortality: results in patients with diastolic blood pressures averaging 115 through 129 mm. Hg. *J. Am. Med. Assoc.*, **202**:1028–1034.

Veterans Administration Cooperative Study Group on Antihypertensive Agents. 1970. II. Results in patients with diastolic blood pressure averaging 90 through 114 mm. Hg. *J. Am. Med. Assoc.*, **213**:1143–1152.