of local recurrence of malignancy after the surgery. Patients were divided into two groups, depending upon the stage of the disease prior to surgery.

Since both the risk and outcome variables are categorical, three measures of treatment effect— difference in recurrence rates, relative risk, and odds ratio—may be computed for each stage (see the calculations in Table 3.2). It turns out that the relative risk is nearly the same for stage 1 and stage 2 patients (3.62 vs. 3.81), whereas the odds ratio and difference in rates depend on the stage (4.02 vs. 5.92 and 0.10 vs. 0.32). In other words, there is an interaction if the treatment effect is expressed in terms of the latter two measures, but no interaction if it is measured by the relative risk.

Since the logarithm of the relative risk is equal to the difference of the log rates ($\log \theta = \log r_1 - \log r_2$), this is an example where an analysis in the original units (recurrence rates) show an interaction, whereas an analysis in a different scale (log − recurrence rates) does not. Often, however, interactions cannot be removed by changing the scale. If in the previous example, stage 1 patients had fewer recurrences with tylectomy than with mastectomy but the opposite had been true for stage 2 patients, there would be no way of avoiding interaction. Figure 3.8 gives another example of nonremovable interaction.

Although it is desirable to avoid interaction since a single measure can then completely describe the treatment effect, sometimes, as we discussed in Section 3.1, because one measure of treatment effect is more useful than others, this measure should be used even if it does result in interaction.

## REFERENCES


Atkins, H., Hayward, J. L., Klugman, D. J., and Wayte, A. B. (1972), Treatment of Early Breast Cancer: A Report after 10 Years of a Clinical Trial, *British Medical Journal,* 423–429.

Berkson, J. (1958), Smoking and Lung Cancer: Some Observations on Two Recent Reports, *Journal of the American Statistical Association,* **53,** 28–38.

Colton, T. (1974), *Statistics in Medicine,* Boston: Little, Brown.

Daniel, D., and Wood, F. S. (1971), *Fitting Equations to Data,* New York: Wiley.

Fleiss, J. S. (1973), *Statistical Methods for Rates and Proportions,* New York: Wiley.

Hanushek, E. A., & Jackson, J. E. (1977), *Statistical Methods for Social Scientists,* New York: Academic Press.

Hill, A. B. (1965), The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine,* **58,** 295–300.

MacMahon, B., and Pugh, T. F. (1970), *Epidemiology, Principles and Methods,* Boston: Little, Brown.

Mosteller, F., and Tukey, J. W. (1977), *Data Analysis and Regression: A Second Course in Statistics,* Reading, MA: Addison-Wesley.

Sheps, M. C. (1959), an Examination of Some Methods of Comparing Rates or Proportions, *Biometrics,* **15,** 87–97.

Tufte, E. R. (1974), *Data Analysis for Politics and Policy,* Englewood Cliffs, N.J.: Prentice-Hall.


CHAPTER 4

# Randomized and Nonrandomized Studies



Estimating a treatment effect requires the construction of a standard of comparison. As we have seen in Chapter 1, this involves a comparison group which does not receive the treatment of interest. In this chapter we will explore several ways of establishing such a comparison group, emphasizing the difference between randomization and other methods. It will be seen that a randomized allocation of subjects to a treatment and control group generally ensures that the latter is an adequate standard of comparison for the former.

We will start by defining randomization and discussing the properties that make this method particularly attractive. We will then give reasons for doing nonrandomized studies, and distinguish the different types of studies involving a comparison group. For simplicity of presentation, this chapter will be confined mainly to studies with a dichotomous risk factor.

## 4.1 DEFINITION OF RANDOMIZATION

*Randomization* is a method whereby subjects are allocated to one of the two risk factor groups by a random mechanism which assures that each individual has an equal chance of being assigned to either group. Tossing a fair coin and allocating an individual to one group or the other based on the appearance of "heads" or the use of a table of random numbers are examples of such processes. For instance, in studying the efficacy of a new medication relative to a standard one, the names of the patients could be entered sequentially, line by line, in a book, and a number from a table of random numbers could be assigned to each line. The patients assigned even numbers would be allocated to the new medication, those with odd numbers to a standard one. A more sophisticated randomized design should be used if we require equal numbers of patients in the two medication groups. After the randomization process has determined that a particular subject should be assigned to a particular group, the investigator must have enough control to implement that assignment. There is clearly no way to conduct a randomized study if the investigator must accept the assignment of people to treatment or comparison groups as determined by nature or by some institutional process (some examples will be given in Sections 4.4 and 4.5).

The primary virtue of randomization is that with high probability the two groups will be similar. Indeed, the only initial systematic difference between the two groups will be that one received the treatment and the other did not. Therefore, if the treatment has no effect, the distribution of the outcome variable in the two groups would be quite similar. In the next section we provide a more extensive discussion of the properties of randomization.

Although randomization offers important advantages, the investigator may sometimes want to consider nonrandom allocations. For example, it is possible to use systematic processes such as allocating every second subject or all the subjects with odd birth years to one of the two groups. Such processes may be much easier to administer than is randomization, and generally these systematic processes will be essentially equivalent to randomization. However, there is always a risk that the characteristic on which the allocation is done (order of arrival, birth year) is related to the outcome under study, so that its effect cannot be disentangled from that of the treatment. Haphazard processes, where no well-defined method is used to form the groups, are even more dangerous since the investigator may allocate, often unconsciously, a particular type of subject to one of the groups. Thus when assignment of subjects to treatments is under the control of the investigator, it is safest to use a random mechanism.

## 4.2 PROPERTIES OF RANDOMIZATION

1. *Randomization generally implies equal distribution of subject characteristics in each group and thereby facilitates causal inference.* If the number of subjects in a randomized study is large, it is unlikely that the two groups differ with respect to any characteristic that can affect the outcome under study, whether or not these characteristics are known to the investigators. To illustrate this property, we consider a hypothetical study to determine whether drug $X$ is effective, as compared to drug $Y$, in reducing blood pressure for patients with hypertension. The investigators identify a number of hypertensive patients. They randomize the patients into a drug $X$ group and a drug $Y$ group. What has been gained by randomization here?



By employing randomization, the investigators assure themselves that the groups are likely to have similar distributions of variables which can affect blood pressure. More precisely, the probability is small that potentially confounding variables differ in the two groups by a large amount. If the drug $X$ group subsequently exhibits a substantially lower average blood pressure than does the drug $Y$ group, randomization makes it unlikely that the difference is caused by a factor other than the drug. For example, it is unlikely that the drug $X$ group has lower blood pressure because it consists of younger people.

2. *Randomization eliminates selection effects.* If individuals found eligible for a study are randomized into groups, there is no possibility that the investigators' initial biases or preferences about which subjects should receive what program could influence the results. For example, in the blood pressure illustration, the investigator, had he or she not randomized, may have tended to give the drug $X$ to the more severe hypertensives. Thus a crude comparison of subsequent blood pressures between the two groups would not give a fair comparison of the two drugs. Or consider the Lanarkshire experiment carried out in schools to study the effect of milk on growth of children. It was criticized by Student (1931) because a loose design allowed the investigators to allocate, perhaps unconsciously, more milk to the poorer and ill-nourished children than to the well-fed children. Although the number of children participating in the study was large, this failure in the design prevented a clear inference about the effect of milk.

If the individuals are considered sequentially for admission to the study, the randomization scheme should be kept secret from the investigator. Otherwise, a medical researcher who knows that the next patient arriving at the hospital will be assigned by the randomization scheme to drug $Y$ may declare that patient ineligible for the study if he or she would have favored drug $X$ for this patient. In this case, randomization together with "blindness" of the investigator will eliminate any selection effects.

The investigator may or may not be conscious of his or her own selectivity in a nonrandomized allocation. Randomization will assure him or her as well as others that subtle selection effects have not operated. This element of persuasiveness is a definite strength of randomized studies.

Selection effects may be created in nonrandomized studies not only by the

investigators, as we have just seen, but by the subjects themselves. With randomization, the subjects cannot select or influence the selection of their own treatment. Self-selection may be particularly troublesome in nonrandomized studies, since it is often difficult to isolate or to measure the variables that distinguish people who select one treatment rather than another, and hence to disentangle the treatment effect from the selection effect. Yerushalmy (1972), in studying the relationship between smoking during pregnancy and birth weight of the infant, argues that the observed difference in birth weights between the smoking and nonsmoking groups may be due to the smoker and not the smoking (i.e., that the smoking may be considered as an index characterizing some other unmeasured differences between smokers and nonsmokers). Even if we know how to measure these differences, it may be impossible to adjust for their effect (see Section 5.6). No such problems arise when randomization is employed.

3. *Randomization provides a basis for statistical inference.* The process of randomization allows us to assign probabilities to observed differences in outcome under the assumption that the treatment has no effect and to perform significance tests (Fisher, 1925). If the significance level attached to an observed difference is very small, it is unlikely that the difference is due only to chance. The purpose of a significance test is to rule out the random explanation. If it is used in conjunction with randomization, it rules out every explanation other than the treatment.

## 4.3 FURTHER POINTS ON RANDOMIZATION

1. *Background variables in randomized studies.* Although the primary virtue of randomization is to tend to balance the two groups with respect to background variables, it does not exclude the possibility of imbalance with respect to one or more individual characteristics. The larger the size of the groups, the less likely this possibility is; however, the investigator should make some basic checks on his or her data to verify that such an unlikely event has not happened. These checks involve comparing the distribution of background variables in the two groups, primarily those background factors which may have an important effect on the outcome factor. If the investigator finds differences between the two groups, he or she should use one of the adjustment techniques described in this book.

The University Group Diabetes Program (UGDP, 1970) provides an example of a carefully randomized study with an extensive check of possible inequalities between treatment groups. The study revealed a higher cardiovascular mortality among patients taking tolbutamide—a drug for the treatment of diabetes, until then regarded as safe—than among patients on other drugs for diabetes or a placebo. One of the controversies that emerged from this study concerned the excessive cardiovascular mortality—12.7% in the tolbutamide group as compared to 6.2% in the placebo group. Could it be explained by an unlucky randomization that happened to assign healthier patients to the control group? A committee appointed by NIH to review the available evidence (Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents, 1975) confirmed that the random process had indeed allocated healthier patients to the control group. After adjusting for this particular problem, however, the committee concluded that there still existed excessive mortality in the tolbutamide group. Cornfield (1971) in his reassessment of the study, points out that when differences in background variables are observed after random allocation, the randomization scheme should be carefully reviewed to eliminate the possibility that it has been violated.



Often, before the randomization is carried out, certain factors are thought to have an important effect on the outcome. It is then advisable to form groups of individuals who are homogeneous in these factors and use randomization within these groups: this process is known as *stratified randomization.* It constitutes an insurance against differences in the distribution of major variables and reduces the random variability. It does require, however, more extensive bookkeeping, to perform the random allocation, and a more complex analysis, to take these groupings into account.

The question of designing such randomized studies and more complex types will not be discussed in this book. The interested reader is referred to texts on experimental design: see Cox (1958) for a nonmathematical presentation and Zelen (1974) and Pocock (1979) for reviews of designs in clinical trials. Kempthorne (1952) and Cochran and Cox (1957) present experimental designs for comparative studies and their analysis at a higher level.

2. *Randomization in small samples.* When the number of individuals in the study is small, the probability of imbalance on important background factors between the groups may be substantial. Precautions should be taken at the design stage to reduce this probability and to decrease the random variation around the difference in outcome. References given in the previous section should be consulted. Also relevant to studies where the individuals are considered sequentially for eligibility—patients entering a hospital, inmates arriving at a prison—are the new type of "biased coin designs" (Efron, 1971; Pocock and Simon, 1975; Simon, 1979). These designs attempt to achieve balance with respect to important background factors while preventing the investigator from being aware of which treatment group the next individual will be assigned to.

## 4.4 REASONS FOR THE USE OF NONRANDOMIZED STUDIES

We have outlined in Section 4.2 three well-known advantages of randomization: (*a*) it tends to balance subject characteristics between the groups and

facilitate causal inference, (*b*) it eliminates selection effects, and (*c*) it provides a basis for statistical inference. Why, then, should standards of comparison be constructed in any other way? We present next some possible reasons for constructing standards of comparison by some other procedure.

1. *Nonrandomized studies are sometimes the only ethical way to conduct an investigation.* If the treatment is potentially harmful, it is generally unethical for an investigator to assign people to this treatment. An example of this is a study of the effects of malnutrition, where we simply cannot assign subjects to intolerable diets. Thus we compare malnourished populations with those on adequate diets.

2. *Nonrandomized studies are sometimes the only ones possible.* Certain investigations require the implementation of treatments that may affect people's lives. In a democratic society randomized implementation of such treatments is not always feasible. Consider, for example, the question of fluoridating a town's water supply. Let us assume that the voters in any town, or their elected representatives, have the final say about whether the water supply is fluoridated. No experimenter can make this decision. We would then have a series of towns, some of which have elected fluoridation and others which have not. The dental experience of the children in these towns can provide a great deal of useful information if properly analyzed.

3. *Nonrandomized studies are usually less expensive.* An advantage of nonrandomized studies is that they usually cost less per subject and may not require the extensive planning and control that are needed for randomized studies. This makes nonrandomized studies particularly attractive in the early stages of any research effort. Preliminary estimates of the relative importance of many background variables and their variation may be developed at a reasonable cost. These data may be important in designing future randomized experiments.

Also, if the investigator is expecting or is interested only in very large, "slam-bang" effects (Gilbert et al., 1975), nonrandomized studies may detect such differences adequately. For instance, the effect of penicillin on mortality was so obvious when it was first used that no randomized study was necessary. However, Gilbert et al. (1975, 1977), after reviewing a large number of innovations in social and medical areas evaluated by randomized and nonrandomized studies, conclude that such slam-bang effects are exceptional.

4. *Nonrandomized studies may be closer to real-life situations.* To the extent that randomization differs from natural selection mechanisms, the conditions of a randomized study might be quite different from those in which the treatment would ordinarily be applied. For example, a program may be very successful for those who choose it themselves on the basis of a media publicity campaign but ineffective when administered as a social experiment. It would then be difficult to disentangle the effect of the program from that of the experimental conditions and to generalize the results of this particular study to a natural, nonexperimental setting. Although we do not discuss this problem of "external validity," because it is primarily subject-matter-related rather than statistical, it might preclude experiments whose conditions of application are too artificial.



## 4.5 TYPES OF COMPARATIVE STUDIES

The investigator who does not have control over the assignment of treatment to individuals can often take advantage of situations created by nature or society. Suppose that we want to study the relation between cigarette smoking (risk factor) and lung cancer (outcome). Although we cannot randomly assign subjects to levels of the risk factor, we can still observe over time groups of people who smoke and who do not and compare the proportions of individuals who develop lung cancer in each group. This approach, called a *cohort study,* may require the observation of a large number of people if the outcome under study is rare in order to get enough "positive outcome" subjects (with lung cancer in this case). In cases of rare outcome, a more economical approach, the *case-control study,* may be considered. One would assemble a group of people with lung cancer and a group without and compare the proportions of smokers in each group.

These two designs (Cochran, 1965; WHO, 1972) can be viewed as different methods of sampling from a given population (we will later refer to that population as the "target population," i.e., the collection of individuals to whom we would like to apply the results of the study). In cohort studies, we focus on risk factor groups and take samples of exposed and unexposed subjects (smokers/nonsmokers); in case-control studies, we focus on outcome groups and take samples of cases and noncases (with lung cancer/without lung cancer). To clarify this point, we can look at a $2 \times 2$ table (Table 4.1) which gives the number of subjects in the target population falling in each category. In a cohort study, we would take samples from the smoking group $(A + B)$ and the nonsmoking group $(C + D)$. In a case-control study, we would take samples from the group with lung cancer $(A + C)$ and the group without lung cancer $(B + D)$.

***Table 4.1 Distribution of Target Population***

| | Smokers | Nonsmokers | Total |
|---|---|---|---|
| With lung cancer | $A$ | $C$ | $A+C$ |
| Without lung cancer | $B$ | $D$ | $B+D$ |
| Total | $A+B$ | $C+D$ | $A+B+C+D$ |

In the following discussion, we will point out arguments for and against each approach—cohort or case-control study. A detailed presentation of these two types of studies may be found in MacMahon and Pugh (1970).

### 4.5.1 Cohort Studies

In a cohort study, persons are selected on the basis of their exposure (or lack of exposure) to the risk factor. The outcome is measured in the subjects of each group after their selection for study.

Cohort studies may be either prospective or retrospective. If the outcome has occurred prior to the start of the study, it is a *retrospective* cohort study. If the outcome has not occurred at the beginning of the study, it is *prospective.* Retrospective studies are particularly useful when the time lag between exposure to the risk factor and the outcome is large, because the time needed to complete a retrospective study is only that needed to assemble and analyze the data. In a prospective study, for instance of smoking and lung cancer, one may have to wait 20 years or more until the risk factor has an effect. The possibility of doing a retrospective cohort study depends on the availability and reliability of records on both the risk and outcome factor. In a prospective study, the investigator can plan and control the collection of data and therefore avoid, or at least be aware of, defects in the collection of data. Note that a randomized study is a special type of prospective cohort study. Consequently, some of the problems mentioned later in this section apply also to randomized studies.

Cohort studies are preferred to case-control studies when the risk factor is rare in the target population. For example, suppose that we want to study the relation between working in a textile mill and lung cancer among all Americans. As this occupation is rather uncommon, it would not be efficient to use a case-control approach, because by selecting individuals with and without lung cancer we would find too small a proportion (if not nonexistent) of cotton textile mill workers to draw any conclusion. (However, if the study were to be restricted to a town with a high proportion of cotton workers, a case-control study might well be appropriate.) But with a cohort study, we could build a sample of these workers which is of reasonable size. Conversely, cohort studies do require very large sample sizes if the outcome is rare.

As pointed out earlier, the latent period between the exposure to the risk factor and the outcome may be very long, so that people may be lost before the outcome is measured. This may happen for several reasons: people move to another region; people do not want to participate any more (e.g., if the study requires periodic measurements); people die (and death is not the outcome under study); and so on. An analysis for handling losses to follow-up is presented in Chapter 11. This method of analysis takes care of situations in which the probability of loss is related to the risk factor. Imagine, for instance, that smokers tend to move more



than nonsmokers and thus the loss rate would be higher in the exposed group than in the nonexposed group. Or people may have an incentive to participate if they belong to one of the risk factor groups: for instance, a tastier diet may be part of a treatment and encourage people to continue to participate; those in the "dull" diet may drop out more easily. More difficulties arise if the probability of loss is related to the outcome factor, as discussed in Chapter 11.

The problem of observation bias in ascertainment of the outcome is often present in cohort studies. If the observer who measures the outcome is aware of the risk factor group to which the subject belongs, he or she may be tempted to set systematically the doubtful cases in one outcome category for the treatment group and in another category for the comparison group. This kind of bias may be avoided by using a "blind" procedure wherein the observer does not know the risk factor group to which the subject belongs. Similarly, it is sometimes possible to keep the subject from knowing to what risk factor group he or she belongs, to avoid differential rates of reporting: the patient who receives a new drug may be influenced by expectations in reporting results of the medication. When both the observer and the subject are kept "blind," the study is called a "double-blind" study.

### 4.5.2 Case-Control Studies

In a case-control study, we assemble groups of subjects on the basis of their outcomes and then collect data on their past exposure to the risk factor. Consider, for example, the following study on the characteristics of adult pedestrians fatally injured by motor vehicles in Manhattan (Haddon et al., 1961). The investigators of this case-control study assembled 50 pedestrians fatally injured by motor vehicles to form the case group and 200 live pedestrians to form the control group. They were interested in the association of different risk factors—age, heavy drinking, and so on—with pedestrian deaths. To ensure some comparability between the two groups, they forced comparability on background variables (other than the risk factors) that they thought related to both the outcome and the risk factors, by assigning to each case four controls with the same sex, found at the same accident site, at the same time of day, and on the same day of week of accident. That is, they "matched" (see Chapter 6) cases and controls on four background variables. Then they compared the age distribution in both groups, the blood alcohol content, and other risk factor distributions in the two groups.

By using a case-control approach, the investigators were able to look at the influence of several risk factors by means of a single study. Their study was rather economical since they needed to assemble data on only 250 people; if they had used a cohort study, tens of thousands of people would have been needed to get a few pedestrian deaths in each risk factor group.

As with all nonrandomized studies, there can be no assurance of comparability between the two groups. The investigators have matched cases and controls on four background factors, but an undetected confounding factor could induce different distributions of risk factors in the two groups even if the risk factors had no effect.

A selection bias may be encountered in case-control studies when cases (or controls) have a different chance of being selected from the target population as exposed than as nonexposed. This results from the fact that the individuals participating in the study are selected after both the exposure and the outcome, so that a combined effect of the risk and outcome may influence the selection. The results of a case-control study have been published (Boston Collaborative Drug Surveillance Program, 1972) showing that the risk of myocardial infarction (MI), more commonly known as heart attack, was twice as large for heavy coffee drinkers as for others. The possibility of a selection bias was later suggested (Maugh, 1973). The cases were selected among hospitalized MI patients and the controls were patients from the same hospitals suffering from other diseases. However, the cases represent a "lucky" fraction of the MI population, since 60% of MI victims die before reaching the hospital. So the data would be consistent with an interpretation that coffee drinking has no adverse effect on the incidence of MI, but on the contrary increases the chance of survival after MI! If that were the case, the proportion of coffee drinkers would be higher among survivors of MI than among MI victims in the target population. Then the cases being sampled from the survivors of MI, rather than from the MI victims in the target population, would have a higher chance of being selected as coffee drinkers than as noncoffee drinkers. We would then find an excess of coffee drinkers among the cases, even though coffee drinking may have no effect on the incidence of MI.

Selection bias makes it hard to generalize the results of a case-control study to a target population, because it may be impossible to know how original groups of exposed and unexposed subjects were reduced by death or migration before they appear as cases or controls (Dorn, 1959; Feinstein, 1973). Important groups of subjects may never appear for observation in a case-control study, such as the 60% of MI victims who die before reaching a hospital.

Observation bias may arise in ascertaining exposure to the risk factor. If it is done by interviews, the quality of memory may be different among cases and controls. For instance, if mothers who give birth to abnormal children are interviewed about a possible exposure to X-rays during their pregnancy, they may remember better all the events that occurred during their pregnancy than will mothers who gave birth to normal children. Also, the interviewer may be inclined to get more accurate information among cases than among controls.

The estimation of effects raises special problems in case-control studies. To illustrate these difficulties, again consider Table 4.1, which classifies subjects



***Table 4.2 Distribution of Individuals in a Hypothetical Case-Control Study***

| | Smokers | Nonsmokers | Total |
|---|---|---|---|
| With lung cancer | 560 | 440 | 1000 |
| Without lung cancer | 360 | 640 | 1000 |

***Table 4.3 Distribution of Individuals in Another Hypothetical Case-Control Study***

| | Smokers | Nonsmokers | Total |
|---|---|---|---|
| With lung cancer | 5,600 | 4,400 | 10,000 |
| Without lung cancer | 360 | 640 | 1,000 |

in the target population according to their smoking and lung cancer statuses. In a case-control study we may decide to take 1000 subjects from the population with lung cancer and 1000 subjects without and look into their smoking history: we may get Table 4.2. Or we may want to choose more lung cancer patients, say 10,000, and get Table 4.3.

In Chapter 3 we discussed three measures of treatment effect available when both the risk and outcome factors are dichotomous. Applied to our example, they would be:

- The difference in lung cancer rates in the smoking and nonsmoking groups.
- The relative risks of developing lung cancer for smokers as compared to nonsmokers.
- The odds ratio, that is, the ratio of the odds of developing lung cancer in the smoking group to that of developing lung cancer in the nonsmoking group.

By comparing Tables 4.2 and 4.3, we see that the lung cancer rate in, for instance, the smoking group, is not meaningful since the number of lung cancer patients may be changed at will by the investigator: in Table 4.2, 560 of 920 (= 560 + 360) smokers have lung cancer; in Table 4.3, 5600 of 5960 (= 5600 + 360) smokers have lung cancer. Thus a measure derived from comparing lung cancer rates, such as the difference of rates and relative risk, cannot be interpreted in a case-control study. Only the odds ratio can be computed, since it does not depend on the sampling ratio of lung cancer to noncancer subjects (it is equal to 2.26 for the data in Tables 4.2 and 4.3).

Under the following two special circumstances, however, the relative risk in the target population may be estimated from the odds ratio of a case-control study:

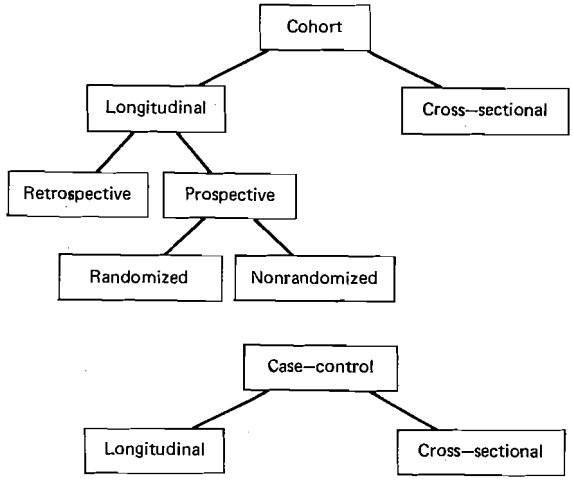
1. The outcome is rare (this implies that the relative risk of the target population is approximately equal to the odds ratio of the target population).
2. There is no selection bias (then the odds ratio of the case-control study is a good estimate of the odds ratio in the target population). Detailed calculations showing how the relative risk in the target population may be estimated by the odds ratio are given in Appendix 4A.

Further information regarding case-control studies can be found in the proceedings of a recent symposium (Ibrahim, 1979).

### 4.5.3 Cross-Sectional Studies

Up to now we have presented two types of studies which are generally *longitudinal;* that is, there is a period between the exposure to the risk factor and the outcome which is the period needed by the risk factor to have an effect, if it has any. However, the length of this period may not be known; the outcome may be undetected for a while or the exposure to the treatment may expand over many years. For these various reasons, *cross-sectional* studies are sometimes done in which the risk and outcome factors are ascertained at the same time. For example, to study the relationship between obesity and heart disease, we might collect data that classified people as obese or nonobese in 1978 and with or

***Table 4.4 Types of Comparative Studies***[a]




[a] All studies are nonrandomized studies except when otherwise indicated. Case-control studies can be carried out only retrospectively.



without heart disease in 1978. In addition to the usual problems in nonrandomized studies there is the difficulty of deciding whether the "outcome" or "risk" factor came first. There is no way in the previous example of indicting obesity as a causal factor in heart disease: we can imagine a circumstance where people with heart disease began to worry about their condition and ate progressively more as their disease worsened.

The different classifications we have made in this chapter are summarized in Table 4.4.

Whereas cohort and case-control studies always involve two different groups to be compared (treated and nontreated for cohort studies, cases and controls for case-control studies), there are instances in which only one group is considered, the group acting as its own comparison group: each subject is measured before and after the treatment, the first measurements providing the comparison, group, the second the treated group. The difficulty of disentangling effects due to the passage of time from the effect of the treatment is particularly troublesome in these studies. The analysis of this type of study is discussed in Chapter 12.

## 4.6 OUR ATTITUDE TOWARD NONRANDOMIZED STUDIES

The limitations of nonrandomized studies that have been discussed in this chapter lead to a question of research strategy. While nonrandomized studies are cheaper, more easily carried out, and can be done retrospectively, inferences from them are generally more suspect than are those from randomized studies. Does this mean that the investigator should discard the idea of doing a study at all if randomization is not feasible? Similarly, when reviewing the results of previous studies, should the reviewer discard all those with nonrandomized designs? We think not. Such a strategy would be extremely conservative.

As mentioned above, there are often sound reasons for considering nonrandomized studies, and much can be learned from them. To eliminate all such studies would be terribly wasteful. On the other hand, the researcher has a responsibility to report clearly all circumstances that may bear on the credibility of results. Without randomization, there are often many alternative explanations of observed results. The researcher must be able to present convincing evidence to rule out alternatives, or to provide data that allow the "consumer" to make an informed judgment.

## APPENDIX 4A THE ODDS RATIO AND THE RELATIVE RISK IN CASE-CONTROL STUDIES

We show how the relative risk in the target population may be estimated from

the odds ratio in a case-control study if conditions 1 and 2 of Section 4.5.2 are satisfied.

Using the notation of Table 4.1, which gives numbers of smokers/nonsmokers with lung cancer/without lung cancer in the target population, we have the relative risk in the target population:

$$\theta = \frac{A/(A+B)}{C/(C+D)} = \frac{A(C+D)}{C(A+B)}.$$

If condition 1 of Section 4.5.2 is satisfied (i.e., lung cancer is rare), then $A$ and $C$ are small compared, respectively, to $A+B$ and $C+D$:

$$A + B \doteq B$$
$$C + D \doteq D,$$

so that $\theta \doteq AD/CB$ = odds ratio ($\psi$) in the target population.

Now denote by lowercase letters the numbers of subjects in a case-control study who represent a sample from the target population. Condition (2) (no selection bias) can be written, ignoring sampling variability, as

$\frac{a}{A} \doteq \frac{c}{C}$ (the selection of lung cancer patients for study does not depend on whether they smoked or not)

$\frac{b}{B} \doteq \frac{d}{D}$ (the selection of control subjects for study does not depend on whether they smoked or not),

so that

$$\frac{A}{C} \doteq \frac{a}{c} \quad \text{and} \quad \frac{D}{B} \doteq \frac{d}{b}.$$

Thus

$$\psi = \frac{AD}{CB} \doteq \frac{ad}{cb} = \hat{\psi} \qquad \text{(the odds ratio in the case-control study)}$$

Therefore, if conditions 1 and 2 are satisfied, $\hat{\psi}$, the odds ratio in a case-control study may be used as an estimate of $\theta$, the relative risk in the target population.

## REFERENCES


Boston Collaborative Drug Surveillance Program (1972), Coffee Drinking and Acute Myocardial Infarction, *Lancet* **2,** 1278–1281.

Cochran, W. G. (1965), The Planning of Observational Studies of Human Populations, *Journal of the Royal Statistical Society, Series A,* **128,** 234–266.

Cochran, W. G., and Cox, G. M. (1957), *Experimental Designs,* 2nd ed., New York: Wiley.





Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents (1975), Report, *Journal of the American Medical Association,* **231,** 583–608.

Cornfield, J. (1971), The University Group Diabetes Program: A Further Statistical Analysis of the Mortality Findings, *Journal of the American Medical Association,* **217,** 1676–1687.

Cox, D. R. (1958), *Planning of Experiments,* New York: Wiley.

Dorn, H. F. (1959), Some Problems Arising in Prospective and Retrospective Studies of the Etiology of Disease, *The New England Journal of Medicine,* **261** (12), 571–579.

Efron, B. (1971), Forcing a Sequential Experiment to Be Balanced, *Biometrika,* **58,** 403–428.

Feinstein, A. R. (1973), The Epidemiologic Trohoc, the Ablative Risk Ratio, and "Retrospective" Research, *Clinical Pharmacology and Therapeutics,* **14,** 291–307.

Fisher, R. A. (1925), *Statistical Methods for Research Workers,* London: Oliver & Boyd.

Gilbert, J. P., Light, R. J., and Mosteller, F. (1975), Assessing Social Innovations: An Empirical Base for Policy, *in* C. A. Bennet and A. A. Lumsdaine, Eds., *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs,* New York: Academic Press.

Gilbert, J. P., McPeek, B., and Mosteller, F. (1977), Progress in Surgery and Anesthesia: Benefits and Risks of Innovative Therapy, *in* J. P. Bunker, B. S. Barnes, and F. Mosteller, Eds., *Costs, Risks and Benefits of Surgery,* New York: Oxford University Press.

Haddon, W., Jr., Valien, P., McCarroll, J. R., and Umberger, C. J. (1961), A Controlled Investigation of the Characteristics of Adult Pedestrians Fatally Injured by Motor Vehicles in Manhattan, *Journal of Chronic Diseases,* **14,** 655–678. Reprinted in E. R. Tufte, Ed., *The Quantitative Analysis of Social Problems,* Reading, MA: Addison-Wesley, 1970, pp. 126–152.

Ibrahim, M. A., Ed. (1979), The Case-Control Study: Consensus and Controversy, *Journal of Chronic Diseases,* **32,** 1–144.

Kempthorne, O. (1952), *The Design and Analysis of Experiments,* New York: Wiley.

MacMahon, B., and Pugh, T. F. (1970), *Epidemiology, Principles and Methods,* Boston: Little, Brown.

Maugh, T. H. (1973), Coffee and Heart Disease: Is There a Link? *Science,* **181,** 534–535.

Pocock, S. J. (1979), Allocation of Patients to Treatment in Clinical Trials, *Biometrics,* **35,** 183–197.

Pocock, S. J., and Simon, R. (1975), Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial, *Biometrics,* **31,** 103–115.

Simon, R. (1979), Restricted Randomization Designs in Clinical Trials, *Biometrics,* **35,** 503–512.

Student (1931), The Lanarkshire Milk Experiment, *Biometrika,* **23,** 398–408.

University Group Diabetes Program (1970), A Study of the Effects of Hypoglycemic Agents on Vascular Complications in Patients with Adult-Onset Diabetes, *Diabetes,* **19** (suppl. 2), 747–783, 787–830.

World Health Organization (1972), Statistical Principles in Public Health Field Studies, *WHO Technical Report 510.*

Yerushalmy, J. (1972), Self-Selection: A Major Problem in Observational Studies, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 4, Berkeley, CA: University of California Press, pp. 329–342.

Zelen, M. (1974), The Randomization and Stratification of Patients to Clinical Trials, *Journal of Chronic Disease,* **27,** 365–375.