CHAPTER 3

Expressing the Treatment Effect

In an ideal hypothetical situation we could observe on the same group of individuals the outcome resulting both from applying and from not applying the treatment. We could then calculate the effect of the treatment by comparing the outcomes under the two conditions. We could define a measure of treatment effect for each individual as the difference between his or her outcomes with and without the treatment. If all subjects were exactly alike, this measure would be the same for each. But more commonly, differences between subjects will cause the measure to vary, possibly in relation to background factors. A treatment may, for instance, be more beneficial to younger than to older people; so the effect would vary with age. We may then wish to define a summary measure of the effect of the treatment on the entire group.

In Section 3.1 we will explore different summary measures of treatment effect. In Example 2.1, Dr. A's choice was to express the treatment effect as the average difference in blood pressure between patients who drink coffee and those who do not drink coffee. We will see that this choice was dictated partly by the nature of the risk factor and partly by the underlying model that Dr. A had in mind as to how coffee consumption affects blood pressure. In Section 3.2 we will leave our ideal situation and see how, when we use a comparison group to estimate a summary measure of treatment effect, a confounding factor may distort that

estimate. Finally, in Section 3.3 we will focus on situations in which the treatment effect is not constant and show that a single summary measure of treatment effect might not be desirable.

3.1 MEASURES OF TREATMENT EFFECT

The choice of measure for treatment effect depends upon the form of the risk and outcome variables. It is useful to make the distinction between a *numerical* variable and a *categorical* variable. The levels of a numerical variable are numbers, whereas the levels of a categorical variable are labels. Thus age expressed in years is a numerical variable, whereas age expressed as young-middle-aged/old or religion expressed as Catholic/Protestant/Jewish/other are categorical variables. Since the levels of a numerical variable are numbers, they can be combined to compute, for instance, a mean (e.g., the mean age of a group of individuals). For categorical variables, on the other hand, the levels are looked at separately (e.g., there are 45 young individuals, 30 middle-aged, and 60 old). Categorical variables with only two possible levels (e.g., intensive reading program vs. standard reading program) are called *dichotomous* variables.

Furthermore, we will sometimes distinguish between an *ordered categorical* variable, such as age, and an *unordered categorical* variable, such as religion. There exists for the first type an intrinsic ordering of the levels (e.g., young/middle-aged/old), whereas for the second type there is no relationship between the levels (e.g., we cannot arrange the various religions in any particular order). A numerical variable can be created from an ordered categorical variable by assigning numbers or *scores* to the different levels (e.g., -1 to young, 0 to middle-aged, and 1 to old).

Using numerical and categorical variables, we can distinguish four different situations, as shown in Figure 3.1. In this book we are concerned mainly with Cases 1 and 2, where the risk variable is categorical.

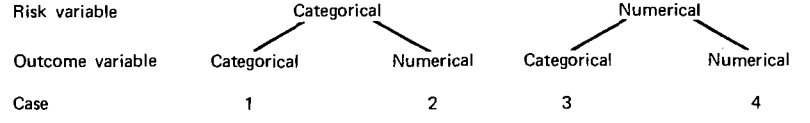


Figure 3.1 *Different cases for measures of treatment effect.*

Case 1: Consider first the effect of a treatment on a dichotomous outcome, specifically death or survival. Three measures of treatment effect are commonly used (Fleiss, 1973; see also Sheps, 1959, for other proposals). We define the three measures and illustrate their use with the data given in Table 3.1. Notice that

Table 3.1 Measures of Treatment Effect for Dichotomous Treatment and Outcome in Three Examples

	Example (a)		Example (b)		Example (c)	
	Treatment	Control	Treatment	Control	Treatment	Control
Death rate	0.06	0.01	0.55	0.50	0.60	0.10
Survival rate	0.94	0.99	0.45	0.50	0.40	0.90
Difference of death rates (Δ)	$0.06 - 0.01 = 0.05$		$0.55 - 0.50 = 0.05$		$0.60 - 0.10 = 0.50$	
Relative risk (θ)	$0.06/0.01 = 6.00$		$0.55/0.50 = 1.10$		$0.60/0.10 = 6.00$	
Odds ratio (ψ)	$\frac{0.06}{0.94}\Big/\frac{0.01}{0.99} = 6.32$		$\frac{0.55}{0.45}\Big/\frac{0.50}{0.50} = 1.22$		$\frac{0.60}{0.40}\Big/\frac{0.10}{0.90} = 13.50$	

in all three examples given in Table 3.1 the treatment is harmful, since the death rate is higher in the treatment group than in the control group. The three measures of treatment effect are:

- The difference in death rates (Δ) between the treatment and control groups. (In epidemiology this is called the attributable risk.) In example (a) in Table 3.1, $\Delta = 0.05$ means that the risk of dying is 0.05 greater in the treatment group.
- The relative risk (θ) is defined as the ratio of the death rate in the treatment group to the death rate in the control group. In example (c) in Table 3.1, $\theta = 6$ implies that the risk of dying in the treatment group (0.60) is 6 times higher than the risk of dying in the control group (0.10).
- The odds ratio (ψ) or *cross-product ratio* is based on the notion of odds. The odds of an event are defined as the ratio of the probability of the event to the probability of its complement. For instance, the odds of dying in the treatment group of example (c) are equal to the death rate (0.60) divided by the survival rate (0.40), or 1.50. When the odds of dying are greater than 1, the risk or probability of dying is greater than that of surviving. Now, the odds ratio in our example is the ratio of the odds of dying in the treatment group (1.50) to the odds of dying in the control group ($0.10/0.90 = 0.11$), or 13.50. The odds of dying are 13.50 times higher in the treatment group. The odds ratio can be conveniently computed as the ratio of the product of the diagonal cells of the treatment by survival table—hence its alternative name, cross-product ratio. In our

example,

$$\psi = \frac{0.60}{0.40}\Big/\frac{0.10}{0.90} \qquad \text{(ratio of the odds)}$$

$$\text{or} = \frac{(0.60) \times (0.90)}{(0.40) \times (0.10)} \qquad \text{(cross-product ratio)}$$

$$= 13.50.$$

The three measures of treatment effect—difference of rates (Δ), relative risk (θ), and odds ratio (ψ)—are linked in the following ways:

1. If the treatment has no effect (i.e., the death rates are equal in the control and treatment groups), then $\Delta = 0$ and $\theta = \psi = 1$.
2. If Δ is negative or θ or ψ smaller than 1, the treatment is beneficial. Conversely, if Δ is positive, θ or ψ greater than 1, the treatment is harmful.
3. If the death rates in the treatment and control groups are low, the odds ratio and relative risk are approximately equal [see, e.g., Table 3.1, example (a); see also Appendix 4A].
4. In certain types of studies (see case-control studies in Chapter 4), only the odds ratio can be meaningfully computed. In these studies the total number of deaths and the total number of survivors are fixed by the investigator, so that death rates and hence differences of death rates and relative risks cannot be interpreted. We shall see in Chapter 4 that the odds ratio does have a sensible interpretation in these studies.

The three examples of Table 3.1 were chosen in such a way that (a) and (b) lead to the same difference of rates and (a) and (c) to the same relative risk. These examples show that the value of one of the three measures has no predictable relation (other than those mentioned above) to the value of any other two: although (a) and (b) have the same Δ of 0.05, their relative risks (6.00 and 1.10) are widely different.

Several factors influence the choice of the measure of treatment effect. The choice may depend on how the measure is going to be used. For example, a difference in death rates would give a better idea of the impact that the treatment would have if it were applied to all diseased people (MacMahon and Pugh, 1970). Berkson (1958; also quoted in Fleiss, 1973), in looking at the effect of smoking on survival, makes this point by saying that "of course, from a strictly practical viewpoint, it is only the total number of increased deaths that matters." On the other hand, the relative risk may highlight a relationship between a risk and an

outcome factor. Hill (1965) remarks that although 71 per 10,000 and 5 per 10,000 are both very low death rates, what "stands out vividly" is that the first is 14 times the second. Thus the choice of a measure may be guided by the aim of the study.

Also, the investigator may believe that one model is more appropriate than another in expressing how the treatment affects the outcome, and he or she can use the data at hand to test his or her belief. That particular model may suggest a measure of treatment effect. This applies for any of the four cases considered in this section. We will turn to Case 2 and illustrate there how a measure may derive from a model.

Case 2: When the outcome variable is numerical (e.g., weight, blood pressure, test score), the difference of the average of the outcome variable between the treatment and comparison groups is a natural measure of treatment effect. For instance, Dr. A can calculate the average blood pressure among coffee drinkers and among non-coffee drinkers and take the difference as a measure of treatment effect.

Dr. A. may think of two different ways in which coffee might affect blood pressure. Let Y_1 and Y_0 be the blood pressure of a given patient with and without coffee drinking. First, coffee drinking might increase blood pressure by a certain amount Δ, which is the same for all patients:

$$Y_1 = Y_0 + \Delta \qquad \text{for any patient (ignoring random variation).}$$

Second, coffee drinking might increase blood pressure proportionally to each patient's blood pressure. Let π be this coefficient of proportionality:

$$Y_1 = \pi Y_0 \qquad \text{for any patient.}$$

By taking logarithms on each side of this expression, we have, equivalently,

$$\log Y_1 = \log Y_0 + \log \pi.$$

Notice that we have transformed a multiplicative effect (π) into an additive effect ($\log \pi$) by changing the scale of the variables through the logarithmic function.

In the first case, Δ would be the measure of treatment effect suggested by the model, which Dr. A. could estimate by the difference of average blood pressure in the coffee and no-coffee group. In the second case, he could consider $\log \pi$ as a measure of treatment effect, which he could estimate by the difference of the average logarithm of blood pressure between the two groups. Or he may find π easier to interpret as a measure of treatment effect and transform back to the original units through the exponential function. Clearly, with the data at hand (see Table 2.1), the first model (and hence Δ) is more appropriate.

Case 3: An example of Case 3, where the risk variable is numerical and the outcome categorical, is a study of increasing doses of a drug on the chance of surviving for 1 year. The odds of dying can be defined for each dose of the drug. The effect of the drug can be assessed by looking at the change in the odds of dying as the dose increases. A model often used in such cases assumes that for any increase of the dose by 1 unit, the logarithm of the odds changes by a constant amount. This amount is taken as the measure of treatment effect.

Case 4: Here both the risk and outcome variables are numerical. Suppose that we want to look at the effect of increasing doses of a drug on blood pressure; if a straight line is fitted to the blood pressure–dose points, the slope of the line can be taken as a measure of the effect of the drug. It represents the change in blood pressure per unit increase in dosage. Regression techniques that can be used in this case will not be discussed in this book. This topic has been covered in many other books (see, e.g., Tufte, 1974; Mosteller and Tukey, 1977; Hanushek and Jackson, 1977; Daniel and Wood, 1971; Colton, 1974).

From the discussion of these four cases, it should be clear that a measure of treatment effect not only depends on the form of the risk and outcome variables, but also on the aim of the study, the scale of the variables, and the models judged appropriate by the investigators.

3.2 WHAT HAPPENS WHEN THERE IS CONFOUNDING

We know from previous chapters that we might be wary of confounding factors when we compare a group of treated individuals and a group of comparison individuals to assess the effect of a treatment. The purpose of this section is to show how a confounding factor distorts the estimate of the treatment effect, and how crude odds ratios or differences of average outcome are not good estimates of treatment effect in the presence of confounding.

As before, we will consider different cases, depending on how the outcome and confounding factors are measured (i.e., whether they are numerical or categorical). We will consider here only dichotomous risk variables, one level being the treatment and the other the comparison. Figure 3.2 illustrates the four possibilities. The numbers (2) and (1) at the top of the figure refer to the case

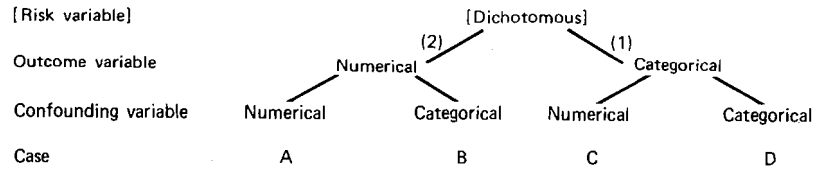


Figure 3.2 *Different cases for the effect of a confounding factor.*

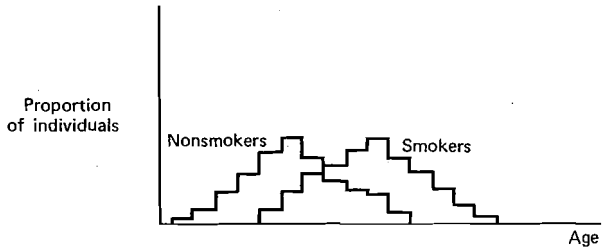


Figure 3.3 *Age distribution in the smoking and nonsmoking groups.*

numbers in Figure 3.1, and these indicate which measures of treatment effect are appropriate for Cases A, B, C, and D.

An example of *Case A* is a study of the effect of smoking on blood pressure where age expressed in years would be a confounding factor. Suppose that the smoking and nonsmoking groups that we compare have the age distributions shown in Figure 3.3. Note that there are very few young smokers and very few old nonsmokers. The average age of smokers is greater than the average age of nonsmokers.

In addition, suppose that a plot of blood pressure vs. age in each group suggests, as in Figure 3.4, that blood pressure is linearly related to age, with equal slopes among smokers and nonsmokers. If we denote blood pressure by Y and age by X and use the subscripts S for smokers and NS for nonsmokers, we have (ignoring random variation)

$$Y_S = \alpha_S + \beta X_S \qquad \text{in the smoking group}$$
$$Y_{NS} = \alpha_{NS} + \beta X_{NS} \qquad \text{in the nonsmoking group.}$$

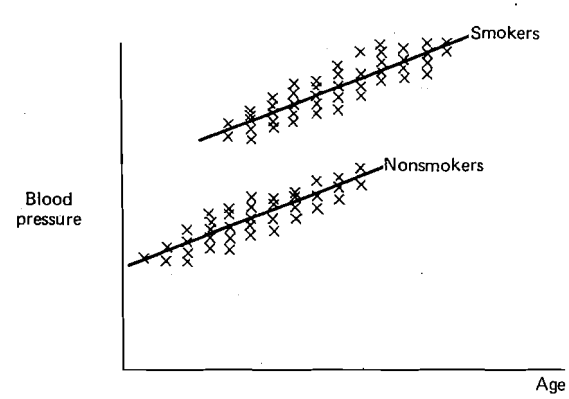


Figure 3.4 *Relationship of blood pressure with age in the smoking and nonsmoking groups.*

The same slope (β) appears in the two equations, but the intercepts α_S and α_{NS} are different.

Note that age satisfies the definition of a confounding factor given in Chapter 2: it has a different distribution in the smoking and nonsmoking groups (Figure 3.3) and it affects blood pressure within each population (Figure 3.4). If we assume that age and smoking are the only factors affecting blood pressure, we can measure the effect of smoking by the vertical distance between the two lines of Figure 3.4 (i.e., $\alpha_S - \alpha_{NS}$).

In the discussion of Case 2 in Section 3.1, we suggested measuring the treatment effect by the difference between the average outcomes: in our example by $\bar{Y}_S - \bar{Y}_{NS}$, the difference between the average blood pressure in the smoking group and that in the nonsmoking group. Since

$$\bar{Y}_S = \alpha_S + \beta \bar{X}_S$$
$$\bar{Y}_{NS} = \alpha_{NS} + \beta \bar{X}_{NS}$$

(where the overbar indicates that we have averaged over the group), it follows that

$$\begin{aligned} \bar{Y}_S - \bar{Y}_{NS} &= (\alpha_S + \beta \bar{X}_S) - (\alpha_{NS} + \beta \bar{X}_{NS}) \\ &= (\alpha_S - \alpha_{NS}) \quad + \beta(\bar{X}_S - \bar{X}_{NS}) \\ &= \text{treatment effect} + \text{bias.} \end{aligned}$$

Thus if we use the difference of average blood pressure, in our example we overestimate the treatment effect by the amount $\beta(\bar{X}_S - \bar{X}_{NS})$, which we call the bias. We have represented this situation in Figure 3.5, which combines Figures 3.3 and 3.4. (In Figure 3.5 the age distributions in each group from Fig.

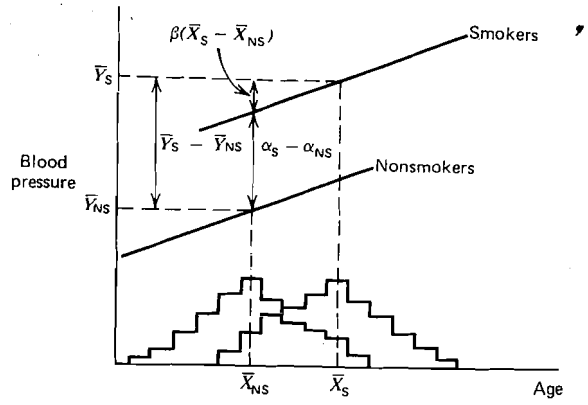


Figure 3.5 *Treatment effect and bias.*

3.3 appear at the bottom of the figure and the relationships between blood pressure and age from Figure 3.4 appear as solid lines. The vertical axis of Figure 3.3 is not explicitly shown.) Note that if age were not a confounding factor, either the age distribution would be the same in the two groups (so that $\bar{X}_S - \bar{X}_{NS} = 0$) or age would not be related to blood pressure (so that $\beta = 0$): in both cases the bias would be 0.

As an example of *Case B,* let us consider sex as a confounding factor. If the difference in mean blood pressures for smokers vs. nonsmokers is the same for males and females, this difference may be regarded as the treatment effect (again assuming that no factors, other than smoking and sex, affect blood pressure). But if males have higher blood pressures than females and if males are more likely to smoke than females, the overall difference in average blood pressure between smokers and nonsmokers is biased as in Case A. Another example of Case B is Example 2.1.

To illustrate *Case C,* where the outcome is categorical and the confounding is numerical, let us suppose that we are interested in the effect of smoking on mortality, and once again we will consider age as a confounding factor. Assume the same age distributions as in the example for Case A (see Figure 3.3). Now consider, for instance, the smoking group: to each level of age corresponds a death rate, and a plot of death rate vs. age may suggest a simple relationship between them; similarly in the nonsmoking group. For instance, in Figure 3.6, we have assumed that the relationship between death rate and age could be described by an exponential curve in each group, or equivalently that the relationship between the logarithm of the death rate and age could be described by a straight line in each group.

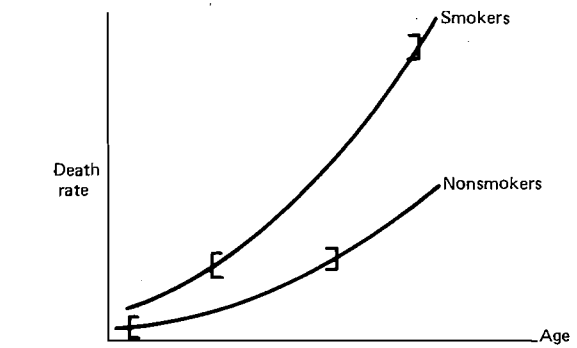
As can be seen in Figure 3.6*b,* we have also assumed that the distance between the straight lines is the same for each age (i.e., the difference in the logarithm of the death rates is a constant $\alpha'_S - \alpha'_{NS}$). Note that this difference is the logarithm of the relative risk, since the relative risk is the ratio of the death rate in the smoking group, r_S, to the death rate in the nonsmoking group, r_{NS}. That is,

$$\log r_S - \log r_{NS} = \alpha'_S - \alpha'_{NS},$$

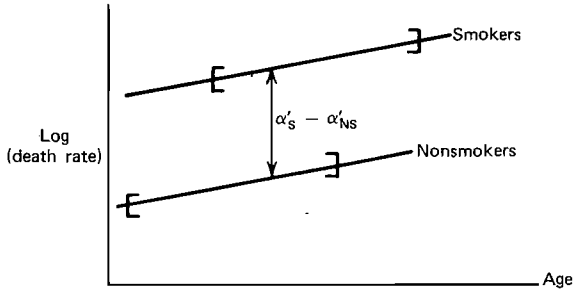
which implies that

$$\log \frac{r_S}{r_{NS}} = \alpha'_S - \alpha'_{NS}.$$

So we are considering a model with the same relative risk at each age. The brackets in Figure 3.6 indicate the ranges of the risks of death for smokers and nonsmokers corresponding to the age ranges of Figure 3.3. A crude relative risk obtained by dividing the overall smoker death rate by the overall nonsmoker death rate would overestimate the true relative risk, because smokers tend to be older than nonsmokers.



(a)



(b)

Figure 3.6 (*a*) *Relationship of death rate with age;* (*b*) *relationship of log* (*death rate*) *with age.*

Confounding in *Case D* operates much the same way as in Case B except that the initial assumption is that the relative risk of death for smokers vs. nonsmokers is the same for males and females. Example 1.1 is of the Case D type.

3.3 TREATMENT EFFECT DEPENDENT ON A BACKGROUND FACTOR

In the previous examples we have assumed an identical treatment effect for all individuals. In Figure 3.4, for instance, smoking increases blood pressure by the same amount for everybody. The assumption of constant treatment effect is commonly made for simplicity, but it may be more realistic to assume that a treatment acts differentially across individuals. This variability may be modeled by assuming that the treatment effect is a function of one or several background factors. For instance, the effect of surgery as compared with standard medication

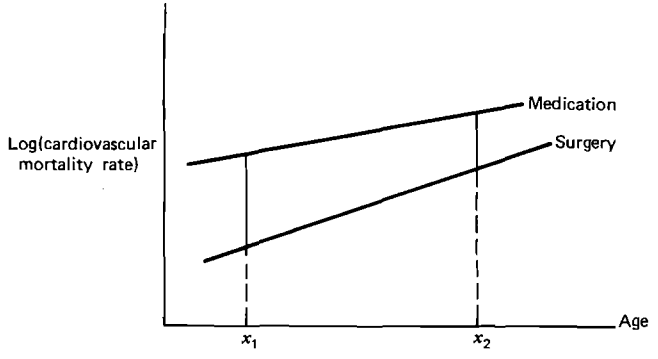


Figure 3.7. *First example of interaction.*

in the treatment of cardiovascular diseases depends in particular on a patient's age, arterial state, and properties of the heart as measured by several variables. It may or may not be desirable to refer then to a summary measure of treatment effect, as the following two hypothetical cases will illustrate.

For simplicity we will assume that the effect of surgery depends only on age and that the relationships between age and cardiovascular mortality for both surgical and medical treatments are as shown in Figures 3.7 and 3.8; in both cases, the logarithm of the cardiovascular mortality rate is a linear function of age under each treatment. In Figure 3.8, but not in Figure 3.7, the two lines cross. In both cases, the comparison of surgery and medication depends on age. In Figure 3.7, surgery is always associated with a lower mortality rate, its greatest benefit being for younger patients (x_1). In this case, a summary treatment effect such as a difference in the average logarithm of the mortality rates would provide useful information on the effect of surgery. Contrast this with Figure 3.8, where surgery is beneficial for younger patients (x_1), whereas for older patients (x_2)

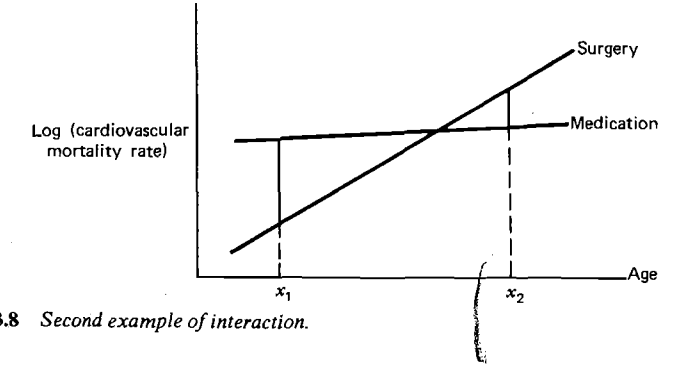


Figure 3.8 *Second example of interaction.*

standard medication is preferable. Here a summary measure would give a distorted picture of the effect of surgery.

When the treatment effect is related to a background factor in this way, there is said to be an *interaction* between the treatment and background factors. The presence or absence of interaction may depend on the measure chosen to express the treatment effect.

Example 3.1 Treatment for breast cancer: Consider the data given in Table 3.2, which come from a randomized study (Atkins et al., 1972) comparing two forms of surgical treatment for breast cancer. The outcome variable is the presence or absence

Table 3.2 Surgical Treatment for Breast Cancer[a]

	Surgical Procedure	
Clinical stage 1	Extended Tylectomy	Radical Mastectomy
Recurrence	15	4
No recurrence	97	104
	112	108
	Rates difference $= \frac{15}{112} - \frac{4}{108} = 0.10$	
	Relative risk $= \frac{15}{112} \Big/ \frac{4}{108} = 3.62$	
	Odds ratio $= \frac{15 \times 104}{4 \times 97} = 4.02$	
Clinical stage 2	Extended Tylectomy	Radical Mastectomy
Recurrence	30	9
No recurrence	40	71
	70	80
	Rates difference $= \frac{30}{70} - \frac{9}{80} = 0.32$	
	Relative risk $= \frac{30}{70} \Big/ \frac{9}{80} = 3.81$	
	Odds ratio $= \frac{30 \times 71}{40 \times 9} = 5.92$	

Adapted, by permission, from Atkins et al. (1972), Tables 2 to 4.

[a] Treatment = surgical procedure; outcome = recurrence; background factor = clinical stage.

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.