

efficacy of taking aspirin by comparing our subsequent headaches. It is more obvious here that “you” and “I” at the same point in time are different units. Your headache status after taking an aspirin can obviously differ from what my headache status would have been had I taken an aspirin. I may be more or less sensitive to aspirin, or I may have started with a more or less severe headache. This type of comparison, often involving many different individuals, is widely used in informal assessments of causal effects, but it is also the basis for many formal studies of causal effects in the social and biomedical sciences. For example, many people view a college education as economically beneficial to future career outcomes based on comparisons of the careers of individuals with, and individuals without, college educations.

By itself, however, the presence of multiple units does not solve the problem of causal inference. Consider the aspirin example with two units, You and I, and two possible treatments for each unit, aspirin or no aspirin. For simplicity, assume that the two available aspirin tablets are equally effective. There are now a total of four treatment levels: you take an aspirin and I do not, I take an aspirin and you do not, we both take an aspirin, or neither of us does. There are therefore four potential outcomes for each of us. For “I” these four potential outcomes are the state of my headache (*i*) if neither of us takes an aspirin, (*ii*) if I take an aspirin and you do not, (*iii*) if you take an aspirin and I do not, and (*iv*) if both of us take an aspirin. “You,” of course, have the corresponding set of four potential outcomes. We can still only observe at most one of these four potential outcomes for each unit, namely the one realized corresponding to whether you and I took, or did not take, an aspirin. Thus each level of the treatment now indicates both whether you take an aspirin and whether I do. In this situation, there are six different comparisons defining causal effects for each of us, depending on which two of the four potential outcomes for each unit are conceptually compared ( $6 = \binom{4}{2}$ ). For example, we can compare the status of my headache if we both take aspirin with the status of my headache if neither of us takes an aspirin, or we can compare the status of my headache if only you take an aspirin to the status of my headache if we both do.

Although we typically make the assumption that whether you take an aspirin does not affect my headache status, it is important to understand the force of such an assumption. One should not lose sight of the fact that it is an assumption, often a strong and controversial one, not a fact, and therefore may be false. Consider a setting where I take aspirin, and I will have a headache if you do not take an aspirin, whereas I will not have a headache if you do take an aspirin: we are in the same room, and unless you take an aspirin to ease your own headache, your incessant complaining will maintain my headache! Such interactions or spillover effects are an important feature of many educational programs, and often motivate changing the unit of analysis from individual children to schools or other groups of individuals.

## 1.6 THE STABLE UNIT TREATMENT VALUE ASSUMPTION

In many situations it may be reasonable to assume that treatments applied to one unit do not affect the outcome for another unit. For example, if we are in different locations and have no contact with each other, it would appear reasonable to assume that whether

you take an aspirin has no effect on the status of my headache. (But, as the example in the previous section illustrates, this assumption need not hold if we are in the same location, and your behavior, itself affected by whether you take an aspirin, may affect the status of my headache, or if we communicate by extrasensory perception.) The stable unit treatment value assumption, or SUTVA (Rubin, 1980a) incorporates both this idea that units do not interfere with one another and the concept that for each unit there is only a single version of each treatment level (ruling out, in this case, that a particular individual could take aspirin tablets of varying efficacy):

### **Assumption 1.1 (SUTVA)**

*The potential outcomes for any unit do not vary with the treatments assigned to other units, and, for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes.*

These two elements of the stability assumption enable us to exploit the presence of multiple units for estimating causal effects.

SUTVA is the first of a number of assumptions discussed in this book that are referred to generally as *exclusion restrictions*: assumptions that rely on external, substantive, information to rule out the existence of a causal effect of a particular treatment relative to an alternative. For instance, in the aspirin example, in order to help make an assessment of the causal effect of aspirin on headaches, we could exclude the possibility that your taking or not taking aspirin has any effect on my headache. Similarly, we could exclude the possibility that the aspirin tablets available to me are of different strengths. Note, however, that these assumptions, and other restrictions discussed later, are not directly informed by observations – they are assumptions. That is, they rely on previously acquired knowledge of the subject matter for their justification. Causal inference is generally impossible without such assumptions, and thus it is critical to be explicit about their content and their justifications.

#### **1.6.1 SUTVA: No Interference**

Consider, first, the no-interference component of SUTVA – the assumption that the treatment applied to one unit does not affect the outcome for other units. Researchers have long been aware of the importance of this concept. For example, when studying the effect of different types of fertilizers in agricultural experiments on plot yields, traditionally researchers have taken care to separate plots using “guard rows,” unfertilized strips of land between fertilized areas. By controlling the leaching of different fertilizers across experimental plots, these guard rows make SUTVA more credible; without them we might suspect that the fertilizer applied to one plot affected the yields in contiguous plots.

In our headache example, in order to address the no-interference assumption, one has to argue, on the basis of a prior knowledge of medicine and physiology, that someone else taking an aspirin in a different location cannot have an effect on my headache. You might think that we could learn about the magnitude of such interference from a separate experiment. Suppose people are paired, with each pair placed in a separate room. In each pair one randomly chosen individual is selected to be the “designated treated” individual and the other the “designated control” individual. Half the pairs are then randomly

selected to be the “treatment pairs” and the other half selected to be “control pairs,” with the “designated treated” individual in the treatment pairs given aspirin and the “designated treated” individual in the control pairs given a placebo. The outcome would then be the status of the headache of the “control” person in each pair. Although such an experiment could shed some light on the plausibility of our no-interference assumption, this experiment relies itself on a more distant version of SUTVA – that treatments assigned to one pair do not affect the results for other pairs. As this example reveals, in order to make any assessment of causal effects, the researcher has to rely on assumed existing knowledge of the current subject matter to assert that some treatments do not affect outcomes for some units.

There exist settings, moreover, in which the no-interference part of SUTVA is controversial. In large-scale job training programs, for example, the outcomes for one individual may well be affected by the number of people trained when that number is sufficiently large to create increased competition for certain jobs. In an extreme example, the effect on your future earnings of going to a graduate program in statistics would surely be very different if everybody your age also went to a graduate program in statistics. Economists refer to this concept as a *general equilibrium* effect, in contrast to a *partial equilibrium* effect, which is the effect on your earnings of a statistics graduate degree under the *ceteris paribus* assumption that “everything else” stayed equal. Another classic example of interference between units arises in settings with immunizations against infectious diseases. The causal effect of your immunization versus no immunization will surely depend on the immunization of others: if everybody else is already immunized with a perfect vaccine, and others can therefore neither get the disease nor transmit it, your immunization is superfluous. However, if no one else is immunized, your treatment (immunization with a perfect vaccine) would be effective relative to no immunization. In such cases, sometimes a more restrictive form of SUTVA can be considered by defining the unit to be the community within which individuals interact, for example, schools in educational settings, or specifically limiting the number of units assigned to a particular treatment.

### 1.6.2 SUTVA: No Hidden Variations of Treatments

The second component of SUTVA requires that an individual receiving a specific treatment level cannot receive different forms of that treatment. Consider again our assessment of the causal effect of aspirin on headaches. For the potential outcome with both of us taking aspirin, we obviously need more than one aspirin tablet. Suppose, however, that one of the tablets is old and no longer contains a fully effective dose, whereas the other is new and at full strength. In that case, each of us may have three treatments available: no aspirin, the ineffective tablet, and the effective tablet. There are thus two forms of the active treatment, both nominally labeled “aspirin”: aspirin+ and aspirin-. Even with no interference we can now think of there being three potential outcomes for each of us, the no aspirin outcome  $Y_i(\text{No Aspirin})$ , the weak aspirin outcome  $Y_i(\text{Aspirin}-)$  and the strong aspirin outcome  $Y_i(\text{Aspirin}+)$ , with  $i$  indexing “I” or “You.” The second part of SUTVA either requires that the two aspirin outcomes are identical:  $Y_i(\text{Aspirin}+) = Y_i(\text{Aspirin}-)$ , or that I can only get Aspirin+ and you can only get Aspirin- (or *vice versa*). Alternatively we can redefine the treatment as taking

a randomly selected aspirin (either Aspirin- or Aspirin+). In that case SUTVA might be satisfied for the redefined stochastic treatment.

Another example of variation in the treatment that is ruled out by SUTVA occurs when differences in the method of administering the treatment matter. The effect of taking a drug for a particular individual may differ depending on whether the individual was assigned to receive it or chose to take it. For example, taking it after being given the choice may lead the individual to take actions that differ from those that would be taken if the individual had no choice in the taking of the drug.

Fundamentally, the second component of SUTVA is again an exclusion restriction. The requirement is that the label of the aspirin tablet, or the nature of the administration of the treatment, cannot alter the potential outcome for any unit. This assumption does *not* require that all forms of each level of the treatment are identical across all units, but only that unit  $i$  exposed to treatment level  $w$  specifies a well-defined potential outcome,  $Y_i(w)$ , for all  $i$  and  $w$ . One strategy to make SUTVA more plausible relies on redefining the represented treatment levels to comprise a larger set of treatments, for example, Aspirin-, Aspirin+, and no-aspirin instead of only Aspirin and no-aspirin. A second strategy involves coarsening the outcome; for example, SUTVA may be more plausible if the outcome is defined to be dead or alive rather than to be a detailed measurement of health status. The point is that SUTVA implies that the potential outcomes for each unit and each treatment are well-defined functions (possibly with stochastic images) of the unit index and the treatment.

### 1.6.3 Alternatives to SUTVA

To summarize the previous discussion, assessing the causal effect of a binary treatment requires observing more than a single unit, because we must have observations of potential outcomes under both treatments: those associated with the receipt of the treatment on some units and those associated with no receipt of it on some other units. However, with more than one unit, we face two immediate complications. First, there exists the possibility that the units interfere with one another, such that one unit's potential outcome when exposed to a specific treatment level, may also depend on the treatment received by another unit. Second, because in multi-unit settings, we must have available more than one copy of each treatment, we may face circumstances in which a unit's potential outcome when receiving the same nominal level of a treatment could vary with different versions of that treatment. These are serious complications, serious in the sense that unless we restrict them by assumptions, combined with careful study design to make these assumptions more realistic, any causal inference will have only limited credibility.

Throughout most of this book, we shall maintain SUTVA. In some cases, however, specific information may suggest that alternative assumptions are more appropriate. For example, in some early AIDS drug trial settings, many patients took some of their assigned drug and shared the remainder with other patients in hopes of avoiding placebos. Given this knowledge, it is clearly no longer appropriate to assert the no-interference element of SUTVA – that treatments assigned to one unit do not affect the outcomes for others. We can, however, use this specific information to model how treatments are received across patients in the study, making alternative – and in this case, more appropriate – assumptions that allow some inference. For example, SUTVA may

be more  
larly, in  
as the u  
exampl  
structu  
is only  
tions b  
many

If we  
plifie  
level  
poten  
man  
ples  
the i  
indi  
trear  
for  
obs

be more appropriate using subgroups of people as units in such AIDS drug trials. Similarly, in educational settings, SUTVA may be more plausible with classrooms or schools as the units of analysis than with students as the units of analysis. In many economic examples, interactions between units are often modeled through assumptions on market structure, again avoiding the no-interference element of SUTVA. Consequently, SUTVA is only one candidate exclusion restriction for modeling the potentially complex interactions between units and the entire set of treatment levels in a particular experiment. In many settings, however, it appears that SUTVA is the leading choice.

## 1.7 THE ASSIGNMENT MECHANISM: AN INTRODUCTION

If we are willing to accept SUTVA, our complicated “You” and “I” aspirin example simplifies to the situation depicted in Table 1.3. Now You and I each face only two treatment levels (e.g., for “You” whether or not “You” take an aspirin), and the accompanying potential outcomes are a function of only our individual actions. This extends readily to many units. To accommodate this generalization, and also the discussion of other examples beyond that of taking or not taking aspirin, as introduced in Section 1.6, let us index the units in the population of size  $N$  by  $i$ , taking on values  $1, \dots, N$ , and let the treatment indicator  $W_i$  take on the values 0 (the control treatment, e.g., no aspirin) and 1 (the active treatment, e.g., aspirin). We have one realized (and possibly observed) potential outcome for each unit. For unit  $i$ , now  $i \in \{1, \dots, N\}$ , let  $Y_i^{\text{obs}}$  denote this realized (and possibly observed) outcome:

$$Y_i^{\text{obs}} = Y_i(W_i) = \begin{cases} Y_i(0) & \text{if } W_i = 0, \\ Y_i(1) & \text{if } W_i = 1. \end{cases}$$

For each unit we also have one missing potential outcome, for unit  $i$  denoted by  $Y_i^{\text{mis}}$ :

$$Y_i^{\text{mis}} = Y_i(1 - W_i) = \begin{cases} Y_i(1) & \text{if } W_i = 0, \\ Y_i(0) & \text{if } W_i = 1. \end{cases}$$

Many writers replace the potential outcomes and treatment indicator with simply the treatment indicator,  $W_i$ , and the observed outcome  $Y_i^{\text{obs}}$ . This “observed-value” notation confuses the objects of inference and the assignment mechanism and can lead to mistakes as we see in Section 1.9.

This information alone, still, does not allow us to infer the causal effect of taking an aspirin on headaches. Suppose, in the two-person headache example, that the person who chose not to take the aspirin did so because he had only a minor headache. Suppose then that an hour later both headaches have faded: the headache for the first person possibly faded because of the aspirin (it would still be there without the aspirin), and the headache of the second person faded simply because it was not a serious headache (it would be gone even without the aspirin). When comparing these two observed potential outcomes, we might conclude that the aspirin had no effect, whereas in fact it may have been the cause of easing the more serious headache. The key piece of information that