Chapter 2

The Experimental Ideal

> It is an important and popular fact that things are not always what they seem. For instance, on the planet Earth, man had always assumed that he was more intelligent than dolphins because he had achieved so much—the wheel, New York, wars and so on—while all the dolphins had ever done was muck about in the water having a good time. But conversely, the dolphins had always believed that they were far more intelligent than man–for precisely the same reasons. In fact there was only one species on the planet more intelligent than dolphins, and they spent a lot of their time in behavioral research laboratories running round inside wheels and conducting frighteningly elegant and subtle experiments on man. The fact that once again man completely misinterpreted this relationship was entirely according to these creatures' plans.
>
> Douglas Adams, *The Hitchhiker's Guide to the Galaxy* (1979)

The most credible and influential research designs use random assignment. A case in point is the Perry preschool project, a 1962 randomized experiment designed to asses the effects of an early-intervention program involving 123 Black preschoolers in Ypsilanti (Michigan). The Perry treatment group was randomly assigned to an intensive intervention that included preschool education and home visits. It's hard to exaggerate the impact of the small but well-designed Perry experiment, which generated follow-up data through 1993 on the participants at age 27. Dozens of academic studies cite or use the Perry findings (see, e.g., Barnett, 1992). Most importantly, the Perry project provided the intellectual basis for the massive Head Start pre-school program, begun in 1964, which ultimately served (and continues to serve) millions of American children.[1]

[1] The Perry data continue to get attention, particular as policy-interest has returned to early education. A recent re-analysis by Michael Anderson (2006) confirms many of the findings from the original Perry study, though Anderson also shows that the overall positive effects of Perry are driven entirely by the impact on girls. The Perry intervention seems to have done nothing for boys.

2.1 The Selection Problem

We take a brief time-out for a more formal discussion of the role experiments play in uncovering causal effects. Suppose you are interested in a causal "if-then" question. To be concrete, consider a simple example: Do hospitals make people healthier? For our purposes, this question is allegorical, but it is surprisingly close to the sort of causal question health economists care about. To make this question more realistic, imagine we're studying a poor elderly population that uses hospital emergency rooms for primary care. Some of these patients are admitted to the hospital. This sort of care is expensive, crowds hospital facilities, and is, perhaps, not very effective (see, e.g., Grumbach, Keane, and Bindman, 1993). In fact, exposure to other sick patients by those who are themselves vulnerable might have a net negative impact on their health.

Since those admitted to the hospital get many valuable services, the answer to the hospital-effectiveness question still seems likely to be "yes". But will the data back this up? The natural approach for an empirically-minded person is to compare the health status of those who have been to the hospital to the health of those who have not. The National Health Interview Survey (NHIS) contains the information needed to make this comparison. Specifically, it includes a question "During the past 12 months, was the respondent a patient in a hospital overnight?" which we can use to identify recent hospital visitors. The NHIS also asks "Would you say your health in general is excellent, very good, good, fair, poor?" The following table displays the mean health status (assigning a 1 to excellent health and a 5 to poor health) among those who have been hospitalized and those who have not (tabulated from the 2005 NHIS):

Group	Sample Size	Mean health status	Std. Error
Hospital	7774	2.79	0.014
No Hospital	90049	2.07	0.003

The difference in the means is 0.71, a large and highly significant contrast in favor of the *non-hospitalized*, with a t-statistic of 58.9.

Taken at face value, this result suggests that going to the hospital makes people sicker. It's not impossible this is the right answer: hospitals are full of other sick people who might infect us, and dangerous machines and chemicals that might hurt us. Still, it's easy to see why this comparison should not be taken at face value: people who go to the hospital are probably less healthy to begin with. Moreover, even after hospitalization people who have sought medical care are not as healthy, on average, as those who never get hospitalized in the first place, though they may well be better than they otherwise would have been.

To describe this problem more precisely, think about hospital treatment as described by a binary random variable, $\mathrm{D}_i = \{0, 1\}$. The outcome of interest, a measure of health status, is denoted by Y_i. The question is whether Y_i is *affected* by hospital care. To address this question, we assume we can imagine what might have happened to someone who went to the hospital if they had not gone and vice versa. Hence, for any individual there are two potential health variables:

$$\textit{potential outcome} = \begin{cases} \mathrm{Y}_{1i} & \text{if } \mathrm{D}_i = 1 \\ \mathrm{Y}_{0i} & \text{if } \mathrm{D}_i = 0 \end{cases}.$$

In other words, Y_{0i} is the health status of an individual had he not gone to the hospital, irrespective of whether he actually went, while Y_{1i} is the individual's health status if he goes. We would like to know the difference between Y_{1i} and Y_{0i}, which can be said to be the causal effect of going to the hospital for individual i. This is what we would measure if we could go back in time and change a person's treatment status.[2]

The observed outcome, Y_i, can be written in terms of potential outcomes as

$$\begin{aligned} \mathrm{Y}_i &= \begin{cases} \mathrm{Y}_{1i} & \text{if } \mathrm{D}_i = 1 \\ \mathrm{Y}_{0i} & \text{if } \mathrm{D}_i = 0 \end{cases} \\ &= \mathrm{Y}_{0i} + (\mathrm{Y}_{1i} - \mathrm{Y}_{0i})\mathrm{D}_i. \end{aligned} \tag{2.1.1}$$

This notation is useful because $\mathrm{Y}_{1i} - \mathrm{Y}_{0i}$ is the causal effect of hospitalization for an individual. In general, there is likely to be a distribution of both Y_{1i} and Y_{0i} in the population, so the treatment effect can be different for different people. But because we never see both potential outcomes for any one person, we must learn about the effects of hospitalization by comparing the average health of those who were and were not hospitalized.

A naive comparison of averages by hospitalization status tells us something about potential outcomes, though not necessarily what we want to know. The comparison of average health conditional on hospitalization status is formally linked to the average causal effect by the equation below:

$$\begin{aligned} \underbrace{E\left[\mathrm{Y}_i|\mathrm{D}_i = 1\right] - E[\mathrm{Y}_i|\mathrm{D}_i = 0]}_{\text{Observed difference in average health}} &= \underbrace{E\left[\mathrm{Y}_{1i}|\mathrm{D}_i = 1\right] - E[\mathrm{Y}_{0i}|\mathrm{D}_i = 1]}_{\text{average treatment effect on the treated}} \\ &\quad + \underbrace{E\left[\mathrm{Y}_{0i}|\mathrm{D}_i = 1\right] - E\left[\mathrm{Y}_{0i}|\mathrm{D}_i = 0\right]}_{\text{selection bias}} \end{aligned}$$

The term

$$E[\mathrm{Y}_{1i}|\mathrm{D}_i = 1] - E[\mathrm{Y}_{0i}|\mathrm{D}_i = 1] = E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i}|\mathrm{D}_i = 1]$$

is the *average causal effect of hospitalization on those who were hospitalized.* This term captures the averages difference between the health of the hospitalized, $E[\mathrm{Y}_{1i}|\mathrm{D}_i = 1]$, and what would have happened to *them* had they not been hospitalized, $E[\mathrm{Y}_{0i}|\mathrm{D}_i = 1]$. The observed difference in health status however, adds to this causal effect a term called *selection bias.* This term is the difference in average Y_{0i} between those who

[2] The potential outcomes idea is a fundamental building block in modern research on causal effects. Important references developing this idea are Rubin (1974, 1977), and Holland (1986), who refers to a causal framework involving potential outcomes as the Rubin Causal Model.

were and were not hospitalized. Because the sick are more likely than the healthy to seek treatment, those who were hospitalized have worse Y_{0i}'s, making selection bias negative in this example. The selection bias may be so large (in absolute value) that it completely masks a positive treatment effect. The goal of most empirical economic research is to overcome selection bias, and therefore to say something about the causal effect of a variable like D_i.

2.2 Random Assignment Solves the Selection Problem

Random assignment of D_i solves the selection problem because random assignment makes D_i independent of potential outcomes. To see this, note that

$$\begin{aligned} E[\text{Y}_i|\text{D}_i = 1] - E[\text{Y}_i|\text{D}_i = 0] &= E[\text{Y}_{1i}|\text{D}_i = 1] - E[\text{Y}_{0i}|\text{D}_i = 0] \\ &= E[\text{Y}_{1i}|\text{D}_i = 1] - E[\text{Y}_{0i}|\text{D}_i = 1], \end{aligned}$$

where the independence of Y_{0i} and D_i allows us to swap $E[\text{Y}_{0i}|\text{D}_i = 1]$ for $E[\text{Y}_{0i}|\text{D}_i = 0]$ in the second line. In fact, given random assignment, this simplifies further to

$$\begin{aligned} E\left[\text{Y}_{1i}|\text{D}_i = 1\right] - E\left[\text{Y}_{0i}|\text{D}_i = 1\right] &= E\left[\text{Y}_{1i} - \text{Y}_{0i}|\text{D}_i = 1\right] \\ &= E\left[\text{Y}_{1i} - \text{Y}_{0i}\right]. \end{aligned}$$

The effect of randomly-assigned hospitalization on the hospitalized is the same as the effect of hospitalization on a randomly chosen patient. The main thing, however, is that random assignment of D_i eliminates selection bias. This does not mean that randomized trials are problem-free, but in principle they solve the most important problem that arises in empirical research.

How relevant is our hospitalization allegory? Experiments often reveal things that are not what they seem on the basis of naive comparisons alone. A recent example from medicine is the evaluation of hormone replacement therapy (HRT). This is a medical intervention that was recommended for middle-aged women to reduce menopausal symptoms. Evidence from the Nurses Health Study, a large and influential non-experimental survey of nurses, showed better health among the HRT users. In contrast, the results of a recently completed randomized trial shows few benefits of HRT. What's worse, the randomized trial revealed serious side effects that were not apparent in the non-experimental data (see, e.g., Women's Health Initiative [WHI], Hsia, *et al.*, 2006).

An iconic example from our own field of labor economics is the evaluation of government-subsidized training programs. These are programs that provide a combination of classroom instruction and on-the-job training for groups of disadvantaged workers such as the long-term unemployed, drug addicts, and ex-offenders. The idea is to increase employment and earnings. Paradoxically, studies based on non-