INTRODUCTION

BLIND MASTER PO: Close your eyes. What do you hear?

YOUNG KWAI CHANG CAINE: I hear the water, I hear the birds.

MASTER PO: Do you hear your own heartbeat?

KWAI CHANG CAINE: No.

MASTER PO: Do you hear the grasshopper that is at your feet?

KWAI CHANG CAINE: Old man, how is it that you hear these things?

MASTER PO: Young man, how is it that you do not?

Kung Fu, Pilot

Economists' reputation for dismality is a bad rap. Economics is as exciting as any science can be: the world is our lab, and the many diverse people in it are our subjects.

The excitement in our work comes from the opportunity to learn about cause and effect in human affairs. The big questions of the day are *our* questions: Will loose monetary policy spark economic growth or just fan the fires of inflation? Iowa farmers and the Federal Reserve chair want to know. Will mandatory health insurance really make Americans healthier? Such policy kindling lights the fires of talk radio. We approach these questions coolly, however, armed not with passion but with data.

Economists' use of data to answer cause-and-effect questions constitutes the field of applied econometrics, known to students and masters alike as *'metrics*. The tools of the 'metrics trade are disciplined data analysis, paired with the machinery of statistical inference. There is a mystical aspect to our work as well: we're after truth, but truth is not revealed in full, and the messages the data transmit require interpretation. In this spirit, we draw inspiration from the journey of Kwai Chang Caine, hero of the classic *Kung Fu* TV series. Caine, a mixed-race Shaolin monk, wanders in search of his U.S.-born half-brother in the nineteenth century American West. As he searches, Caine questions all he sees in human affairs, uncovering hidden relationships and deeper meanings. Like Caine's journey, the Way of 'Metrics is illuminated by questions.

Other Things Equal

In a disturbing development you may have heard of, the proportion of American college students completing their degrees in a timely fashion has taken a sharp turn south. Politicians and policy analysts blame falling college graduation rates on a pernicious combination of tuition hikes and the large student loans many students use to finance their studies. Perhaps increased student borrowing derails some who would otherwise stay on track. The fact that the students most likely to drop out of school often shoulder large student loans would seem to substantiate this hypothesis.

You'd rather pay for school with inherited riches than borrowed money if you can. As we'll discuss in detail, however, education probably boosts earnings enough to make loan repayment bearable for most graduates. How then should we interpret the negative correlation between debt burden and college graduation rates? Does indebtedness *cause* debtors to drop out? The first question to ask in this context is who borrows the most. Students who borrow heavily typically come from middle and lower income families, since richer families have more savings. For many reasons, students from lower income families are less likely to complete a degree than those from higher income families, regardless of whether they've borrowed heavily. We should therefore be skeptical of claims that high debt burdens cause lower college completion rates when these claims are based solely on comparisons of completion rates between those with more or less debt. By virtue of the correlation between family background and college debt, the contrast in graduation rates between those with and without student loans is not an *other things equal* comparison.

As college students majoring in economics, we first learned the *other things equal* idea by its Latin name, *ceteris paribus*. Comparisons made under *ceteris paribus* conditions have a causal interpretation. Imagine two students identical in every way, so their families have the same financial resources and their parents are similarly educated.

One of these virtual twins finances college by borrowing and the other from savings. Because they are otherwise equal in every way (their grandmother has treated both to a small nest egg), differences in their educational attainment can be attributed to the fact that only one has borrowed. To this day, we wonder why so many economics students first encounter this central idea in Latin; maybe it's a conspiracy to keep them from thinking about it. Because, as this hypothetical comparison suggests, real *other things equal* comparisons are hard to engineer, some would even say *impossibile* (that's Italian not Latin, but at least people still speak it).

Hard to engineer, maybe, but not necessarily impossible. The 'metrics craft uses data to get to *other things equal* in spite of the obstacles—called selection bias or omitted variables bias—found on the path running from raw numbers to reliable causal knowledge. The path to causal understanding is rough and shadowed as it snakes around the boulders of selection bias. And yet, masters of 'metrics walk this path with confidence as well as humility, successfully linking cause and effect.

Our first line of attack on the causality problem is a randomized experiment, often called a randomized trial. In a randomized trial, researchers change the causal variables of interest (say, the availability of college financial aid) for a group selected using something like a coin toss. By changing circumstances randomly, we make it highly likely that the variable of interest is unrelated to the many other factors determining the outcomes we mean to study. Random assignment isn't the same as holding everything else fixed, but it has the same effect. Random manipulation makes *other things equal* hold on average across the groups that did and did not experience manipulation. As we explain in Chapter 1, "on average" is usually good enough.

Randomized trials take pride of place in our 'metrics toolkit. Alas, randomized social experiments are expensive to field and may be slow to bear fruit, while research funds are scarce and life is short. Often, therefore, masters of 'metrics turn to less powerful but more accessible research designs. Even when we can't practicably randomize, however, we still dream of the trials we'd like to do. The notion of an ideal experiment disciplines our approach to econometric research. *Mastering 'Metrics* shows how wise application of our five favorite econometric tools brings us as close as possible to the causality-revealing power of a real experiment.

Our favorite econometric tools are illustrated here through a series of well-crafted and important econometric studies. Vetted by Grand Master Oogway of *Kung Fu Panda*'s Jade Palace, these investigations of causal effects are distinguished by their awesomeness. The methods they use—random assignment, regression, instrumental variables, regression discontinuity designs, and differences-in-differences—are the Furious Five of econometric research. For starters, motivated by the contemporary American debate over health care, the first chapter describes two social experiments that reveal whether, as many policymakers believe, health insurance indeed helps those who have it stay healthy. Chapters 2–5 put our other tools to work, crafting answers to important questions ranging from the benefits of attending private colleges and selective high schools to the costs of teen drinking and the effects of central bank injections of liquidity.

Our final chapter puts the Furious Five to the test by returning to the education arena. On average, college graduates earn about twice as much as high school graduates, an earnings gap that only seems to be growing. Chapter 6 asks whether this gap is evidence of a large causal return to schooling or merely a reflection of the many other advantages those with more education might have (such as more educated parents). Can the relationship between schooling and earnings ever be evaluated on a *ceteris paribus* basis, or must the boulders of selection bias forever block our way? The challenge of quantifying the causal link between schooling and earnings provides a gripping test match for 'metrics tools and the masters who wield them.

Mastering 'Metrics

Chapter 1

Randomized Trials

KWAI CHANG CAINE: What happens in a man's life is already written. A man must move through life as his destiny wills.

OLD MAN: Yet each is free to live as he chooses. Though they seem opposite, both are true.

Kung Fu, Pilot

Our Path

Our path begins with experimental *random assignment,* both as a framework for causal questions and a benchmark by which the results from other methods are judged. We illustrate the awesome power of random assignment through two randomized evaluations of the effects of health insurance. The appendix to this chapter also uses the experimental framework to review the concepts and methods of statistical inference.

1.1 In Sickness and in Health (Insurance)

The Affordable Care Act (ACA) has proven to be one of the most controversial and interesting policy innovations we've seen. The ACA requires Americans to buy health insurance, with a tax penalty for those who don't voluntarily buy in. The question of the proper role of government in the market for health care has many angles. One is the causal effect of health insurance on health. The United States spends more of its GDP on health care than do other developed nations, yet Americans are surprisingly unhealthy. For example, Americans are more likely to be overweight and die sooner than their Canadian cousins, who spend only about two-thirds as much on care. America is also unusual among developed countries in having no universal health insurance scheme. Perhaps there's a causal connection here.

Elderly Americans are covered by a federal program called Medicare, while some poor Americans (including most single mothers, their children, and many other poor children) are covered by Medicaid. Many of the working, prime-age poor, however, have long been uninsured. In fact, many uninsured Americans have chosen not to participate in an employer-provided insurance plan.[1] These workers, perhaps correctly, count on hospital emergency departments, which cannot turn them away, to address their health-care needs. But the emergency department might not be the best place to treat, say, the flu, or to manage chronic conditions like diabetes and hypertension that are so pervasive among poor Americans. The emergency department is not required to provide long-term care. It therefore stands to reason that government-mandated health insurance might yield a health dividend. The push for subsidized universal health insurance stems in part from the belief that it does.

The *ceteris paribus* question in this context contrasts the health of someone with insurance coverage to the health of the same person were they without insurance (other than an emergency department backstop). This contrast highlights a fundamental empirical conundrum: people are either insured or not. We don't get to see them both ways, at least not at the same time in exactly the same circumstances.

In his celebrated poem, "The Road Not Taken," Robert Frost used the metaphor of a crossroads to describe the causal effects of personal choice:

Two roads diverged in a yellow wood,

And sorry I could not travel both

And be one traveler, long I stood

And looked down one as far as I could

To where it bent in the undergrowth;

Frost's traveler concludes:

Two roads diverged in a wood, and I—

I took the one less traveled by,

And that has made all the difference.

The traveler claims his choice has mattered, but, being only one person, he can't be sure. A later trip or a report by other travelers won't nail it down for him, either. Our narrator might be older and wiser the second time around, while other travelers might have different experiences on the same road. So it is with any choice, including those related to health insurance: would uninsured men with heart disease be disease-free if they had insurance? In the novel *Light Years,* James Salter's irresolute narrator observes: "Acts demolish their alternatives, that is the paradox." We can't know what lies at the end of the road not taken.

We can't know, but evidence can be brought to bear on the question. This chapter takes you through some of the evidence related to paths involving health insurance. The starting point is the National Health Interview Survey (NHIS), an annual survey of the U.S. population with detailed information on health and health insurance. Among many other things, the NHIS asks: "Would you say your health in general is excellent, very good, good, fair, or poor?" We used this question to code an index that assigns 5 to excellent health and 1 to poor health in a sample of married 2009 NHIS respondents who may or may not be insured.[2] This index is our *outcome:* a measure we're interested in studying. The causal relation of interest here is determined by a variable that indicates coverage by private health insurance. We call this variable the *treatment,* borrowing from the literature on medical trials, although the treatments we're interested in need not be medical treatments like drugs or surgery. In this context, those with insurance can be thought of as the *treatment group;* those without insurance make up the *comparison* or *control group.* A good control group reveals the fate of the treated in a counterfactual world where they are not treated.

The first row of Table 1.1 compares the average health index of insured and uninsured Americans, with statistics tabulated separately for husbands and wives.[3] Those with health insurance are indeed healthier than those without, a gap of about .3 in the index for men and .4 in the index for women. These are large differences when measured against the standard deviation of the health index, which is about 1. (Standard deviations, reported in square brackets in Table

1.1, measure variability in data. The chapter appendix reviews the relevant formula.) These large gaps might be the health dividend we're looking for.

Fruitless and Fruitful Comparisons

Simple comparisons, such as those at the top of Table 1.1, are often cited as evidence of causal effects. More often than not, however, such comparisons are misleading. Once again the problem is *other things equal,* or lack thereof. Comparisons of people with and without health insurance are not apples to apples; such contrasts are apples to oranges, or worse.

Among other differences, those with health insurance are better educated, have higher income, and are more likely to be working than the uninsured. This can be seen in panel B of Table 1.1, which reports the average characteristics of NHIS respondents who do and don't have health insurance. Many of the differences in the table are large (for example, a nearly 3-year schooling gap); most are statistically precise enough to rule out the hypothesis that these discrepancies are merely chance findings (see the chapter appendix for a refresher on statistical significance). It won't surprise you to learn that most variables tabulated here are highly correlated with health as well as with health insurance status. More-educated people, for example, tend to be healthier as well as being overrepresented in the insured group. This may be because more-educated people exercise more, smoke less, and are more likely to wear seat belts. It stands to reason that the difference in health between insured and uninsured NHIS respondents at least partly reflects the extra schooling of the insured.

TABLE 1.1

Health and demographic characteristics of insured and uninsured couples in the NHIS

	Husbands			Wives		
	Some HI (1)	No HI (2)	Difference (3)	Some HI (4)	No HI (5)	Difference (6)
			A. Health			
Health index	4.01 [.93]	3.70 [1.01]	.31 (.03)	4.02 [.92]	3.62 [1.01]	.39 (.04)
			B. Characteristics			
Nonwhite	.16	.17	−.01 (.01)	.15	.17	−.02 (.01)
Age	43.98	41.26	2.71 (.29)	42.24	39.62	2.62 (.30)
Education	14.31	11.56	2.74 (.10)	14.44	11.80	2.64 (.11)
Family size	3.50	3.98	−.47 (.05)	3.49	3.93	−.43 (.05)
Employed	.92	.85	.07 (.01)	.77	.56	.21 (.02)
Family income	106,467	45,656	60,810 (1,355)	106,212	46,385	59,828 (1,406)
Sample size	8,114	1,281		8,264	1,131	

Notes: This table reports average characteristics for insured and uninsured married couples in the 2009 National Health Interview Survey (NHIS). Columns (1), (2), (4), and (5) show average characteristics of the group of individuals specified by the column heading. Columns (3) and (6) report the difference between the average characteristic for individuals with and without health insurance (HI). Standard deviations are in brackets; standard errors are reported in parentheses.

Our effort to understand the causal connection between insurance and health is aided by fleshing out Frost's two-roads metaphor. We use the letter Y as shorthand for health, the outcome variable of interest. To make it clear when we're talking about specific people, we use subscripts as a stand-in for names: Y_i is the health of individual *i*. The outcome Y_i is recorded in our data. But, facing the choice of whether to pay for health insurance, person *i* has two *potential outcomes,* only one of which is observed. To distinguish one potential outcome from another, we add a second subscript: The road taken without health insurance leads to Y_{0i} (read this as "y-zero-i") for person *i*, while the road with health insurance leads to Y_{1i} (read this as "y-one–i") for person *i*. Potential outcomes lie at the end of each road one *might* take. The causal effect of insurance on health is the difference between them, written $Y_{1i} - Y_{0i}$.[4]

To nail this down further, consider the story of visiting Massachusetts Institute of Technology (MIT) student Khuzdar Khalat, recently arrived from Kazakhstan. Kazakhstan has a national health insurance system that covers all its citizens automatically (though you wouldn't go there just for the health insurance). Arriving in Cambridge, Massachusetts, Khuzdar is surprised to learn that MIT students must decide whether to opt in to the university's health insurance plan, for which MIT levies a hefty fee. Upon reflection, Khuzdar judges the MIT insurance worth paying for, since he fears upper respiratory infections in chilly New England. Let's say that $Y_{0i} = 3$ and $Y_{1i} = 4$ for i = Khuzdar. For him, the causal effect of insurance is one step up on the NHIS scale:

$$Y_{1,\text{Khuzdar}} - Y_{0,\text{Khuzdar}} = 1.$$

Table 1.2 summarizes this information.

TABLE 1.2
Outcomes and treatments for Khuzdar and Maria

	Khuzdar Khalat	Maria Moreño
Potential outcome without insurance: Y_{0i}	3	5
Potential outcome with insurance: Y_{1i}	4	5
Treatment (insurance status chosen): D_i	1	0
Actual health outcome: Y_i	4	5
Treatment effect: $Y_{1i} - Y_{0i}$	1	0

It's worth emphasizing that Table 1.2 is an imaginary table: some of the information it describes must remain hidden. Khuzdar will either buy insurance, revealing his value of Y_{1i}, or he won't, in which case his Y_{0i} is revealed. Khuzdar has walked many a long and dusty road in Kazakhstan, but even he cannot be sure what lies at the end of those not taken.

Maria Moreño is also coming to MIT this year; she hails from Chile's Andean highlands. Little concerned by Boston winters, hearty Maria is not the type to fall sick easily. She therefore passes up the MIT insurance, planning to use her money for travel instead. Because Maria has $Y_{0,\text{Maria}} = Y_{1,\text{Maria}} = 5$, the causal effect of insurance on her health is

$$Y_{1,\text{Maria}} - Y_{0,\text{Maria}} = 0.$$

Maria's numbers likewise appear in Table 1.2.

Since Khuzdar and Maria make different insurance choices, they offer an interesting comparison. Khuzdar's health is $Y_{\text{Khuzdar}} = Y_{1,\text{Khuzdar}} = 4$, while Maria's is $Y_{\text{Maria}} = Y_{0,\text{Maria}} = 5$. The difference between them is

$$Y_{\text{Khuzdar}} - Y_{\text{Maria}} = -1.$$

Taken at face value, this quantity—which we observe—suggests Khuzdar's decision to buy insurance is counterproductive. His MIT insurance coverage notwithstanding, insured Khuzdar's health is worse than uninsured Maria's.

In fact, the comparison between frail Khuzdar and hearty Maria tells us little about the causal effects of their choices. This can be seen by linking observed and potential outcomes as follows:

$$\begin{aligned} Y_{\text{Khuzdar}} - Y_{\text{Maria}} &= Y_{1,\text{Khuzdar}} - Y_{0,\text{Maria}} \\ &= \underbrace{Y_{1,\text{Khuzdar}} - Y_{0,\text{Khuzdar}}}_{1} + \underbrace{\{Y_{0,\text{Khuzdar}} - Y_{0,\text{Maria}}\}}_{-2}. \end{aligned}$$

The second line in this equation is derived by adding and subtracting $Y_{0,\text{Khuzdar}}$, thereby generating two hidden comparisons that determine the one we see. The first comparison, $Y_{1,\text{Khuzdar}} - Y_{0,\text{Khuzdar}}$, is the causal effect of health insurance on Khuzdar, which is equal to 1. The second, $Y_{0,\text{Khuzdar}} - Y_{0,\text{Maria}}$, is the difference between the two students' health status were both to decide against insurance. This term, equal to -2, reflects Khuzdar's relative frailty. In the context of our effort to uncover causal effects, the lack of comparability captured by the second term is called *selection bias.*

You might think that selection bias has something to do with our focus on particular individuals instead of on groups, where, perhaps, extraneous differences can be expected to "average out." But the difficult problem of selection bias carries over to comparisons of groups, though, instead of individual causal effects, our attention shifts to *average causal effects.* In a group of n people, average causal effects are written $Avg_n[Y_{1i} - Y_{0i}]$, where averaging is done in the usual way (that is, we sum individual outcomes and divide by n):

$$\begin{aligned} Avg_n[Y_{1i} - Y_{0i}] &= \frac{1}{n}\sum_{i=1}^{n}[Y_{1i} - Y_{0i}] \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_{1i} - \frac{1}{n}\sum_{i=1}^{n} Y_{0i}. \qquad (1.1) \end{aligned}$$

The symbol $\sum_{i=1}^{n}$ indicates a sum over everyone from $i = 1$ to n, where n is the size of the group over which we are averaging. Note that both summations in equation (1.1) are taken over everybody in the group of interest. The average causal effect of health insurance compares average health in hypothetical scenarios where everybody in the group does and does not have health insurance. As a computational matter, this is the average of individual causal effects like $Y_{1,\text{Khuzdar}} - Y_{0,\text{Khuzdar}}$ and $Y_{1,\text{Maria}} - Y_{0,\text{Maria}}$ for each student in our data.

An investigation of the average causal effect of insurance naturally begins by comparing the average health of groups of insured and uninsured people, as in Table 1.1. This comparison is facilitated by the construction of a *dummy variable,* D_i, which takes on the values 0 and 1 to indicate insurance status:

$$D_i = \begin{cases} 1 & \text{if } i \text{ is insured} \\ 0 & \text{otherwise.} \end{cases}$$

We can now write $Avg_n[Y_i|D_i = 1]$ for the average among the insured and $Avg_n[Y_i|D_i = 0]$ for the average among the uninsured. These quantities are averages *conditional* on insurance status.[5]

The average Y_i for the insured is necessarily an average of outcome Y_{1i}, but contains no information about Y_{0i}. Likewise, the average Y_i

among the uninsured is an average of outcome Y_{0i}, but this average is devoid of information about the corresponding Y_{1i}. In other words, the road taken by those with insurance ends with Y_{1i}, while the road taken by those without insurance leads to Y_{0i}. This in turn leads to a simple but important conclusion about the difference in average health by insurance status:

$$\begin{aligned}&\textit{Difference in group means}\\&\quad = Avg_n[Y_i|D_i = 1] - Avg_n[Y_i|D_i = 0]\\&\quad = Avg_n[Y_{1i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0], \quad (1.2)\end{aligned}$$

an expression highlighting the fact that the comparisons in Table 1.1 tell us something about potential outcomes, though not necessarily what we want to know. We're after $Avg_n[Y_{1i} - Y_{0i}]$, an average causal effect involving everyone's Y_{1i} and everyone's Y_{0i}, but we see average Y_{1i} only for the insured and average Y_{0i} only for the uninsured.

To sharpen our understanding of equation (1.2), it helps to imagine that health insurance makes everyone healthier by a constant amount, κ. As is the custom among our people, we use Greek letters to label such *parameters,* so as to distinguish them from variables or data; this one is the letter "kappa." The *constant-effects assumption* allows us to write:

$$Y_{1i} = Y_{0i} + \kappa, \qquad (1.3)$$

or, equivalently, $Y_{1i} - Y_{0i} = \kappa$. In other words, κ is both the individual and average causal effect of insurance on health. The question at hand is how comparisons such as those at the top of Table 1.1 relate to κ.

Using the constant-effects model (equation (1.3)) to substitute for $Avg_n[Y_{1i}|D_i = 1]$ in equation (1.2), we have:

$$\begin{aligned}&Avg_n[Y_{1i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0]\\&\quad = \{\kappa + Avg_n[Y_{0i}|D_i = 1]\} - Avg_n[Y_{0i}|D_i = 0]\\&\quad = \kappa + \{Avg_n[Y_{0i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0]\}. \quad (1.4)\end{aligned}$$

This equation reveals that health comparisons between those with and without insurance equal the causal effect of interest (κ) plus the difference in average Y_{0i} between the insured and the uninsured. As in the parable of Khuzdar and Maria, this second term describes selection bias. Specifically, the difference in average health by insurance status can be written:

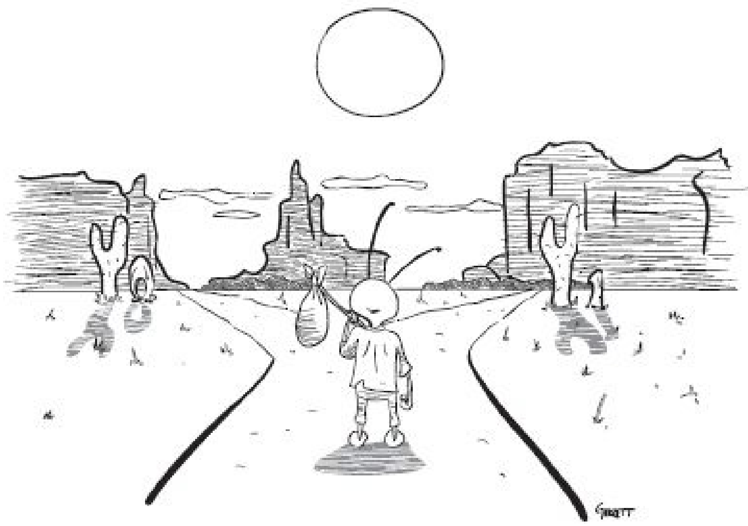
$$\begin{aligned}&\textit{Difference in group means}\\&\quad = \textit{Average causal effect} + \textit{Selection bias},\end{aligned}$$

where selection bias is defined as the difference in average Y_{0i} between the groups being compared.

How do we know that the difference in means by insurance status is contaminated by selection bias? We know because Y_{0i} is shorthand for everything about person i related to health, other than insurance status. The lower part of Table 1.1 documents important noninsurance differences between the insured and uninsured, showing that *ceteris* isn't *paribus* here in many ways. The insured in the NHIS are healthier for all sorts of reasons, including, perhaps, the causal effects of insurance. But the insured are also healthier because they are more educated, among other things. To see why this matters, imagine a world in which the causal effect of insurance is zero (that is, $\kappa = 0$). Even in such a world, we should expect insured NHIS respondents to be healthier, simply because they are more educated, richer, and so on.

We wrap up this discussion by pointing out the subtle role played by information like that reported in panel B of Table 1.1. This panel shows that the groups being compared differ in ways that we can observe. As we'll see in the next chapter, if the only source of selection bias is a set of differences in characteristics that we can observe and measure, selection bias is (relatively) easy to fix. Suppose, for example, that the only source of selection bias in the insurance comparison is education. This bias is eliminated by focusing on samples of people with the same schooling, say, college graduates. Education is the same for insured and uninsured people in such a sample, because it's the same for everyone in the sample.

The subtlety in Table 1.1 arises because when observed differences proliferate, so should our suspicions about unobserved differences. The fact that people with and without health insurance differ in many visible ways suggests that even were we to hold observed characteristics fixed, the uninsured would likely differ from the insured in ways we don't see (after all, the list of variables we can see is partly fortuitous). In other words, even in a sample consisting of insured and uninsured people with the same education, income, and employment status, the insured might have higher values of Y_{0i}. The principal challenge facing masters of 'metrics is elimination of the selection bias that arises from such unobserved differences.



Breaking the Deadlock: Just RANDomize

> My doctor gave me 6 months to live … but when I couldn't pay the bill, he gave me 6 months more.
>
> Walter Matthau

Experimental random assignment eliminates selection bias. The logistics of a randomized experiment, sometimes called a *randomized trial,* can be complex, but the logic is simple. To study the effects of health insurance in a randomized trial, we'd start with a sample of people who are currently uninsured. We'd then provide health insurance to a randomly chosen subset of this sample, and let the rest go to the emergency department if the need arises. Later, the health of the insured and uninsured groups can be compared. Random assignment makes this comparison *ceteris paribus*: groups insured and uninsured by random assignment differ only in their insurance status