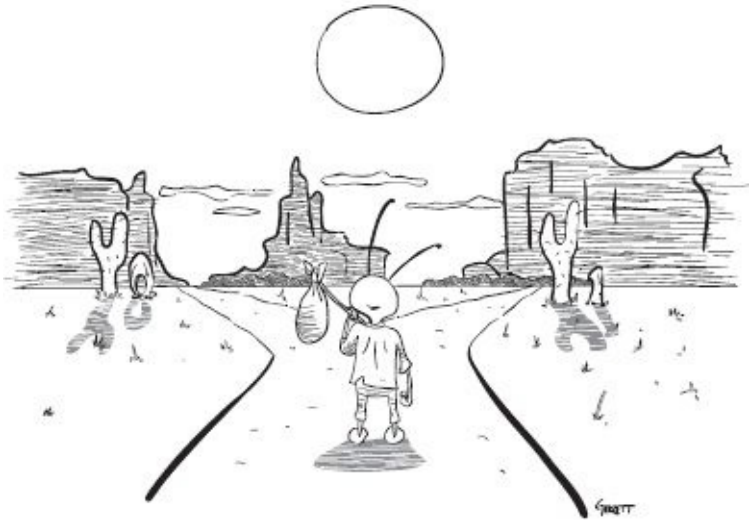
proliferate, so should our suspicions about unobserved differences. The fact that people with and without health insurance differ in many visible ways suggests that even were we to hold observed characteristics fixed, the uninsured would likely differ from the insured in ways we don't see (after all, 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

and any consequences that follow from it.

Suppose the MIT Health Service elects to forgo payment and tosses a coin to determine the insurance status of new students Ashish and Zandile (just this once, as a favor to their distinguished Economics Department). Zandile is insured if the toss comes up heads; otherwise, Ashish gets the coverage. A good start, but not good enough, since random assignment of two experimental subjects does not produce insured and uninsured apples. For one thing, Ashish is male and Zandile female. Women, as a rule, are healthier than men. If Zandile winds up healthier, it might be due to her good luck in having been born a woman and unrelated to her lucky draw in the insurance lottery. The problem here is that two is not enough to tango when it comes to random assignment. We must randomly assign treatment in a sample that's large enough to ensure that differences in individual characteristics like sex wash out.

Two randomly chosen groups, when large enough, are indeed comparable. This fact is due to a powerful statistical property known as the *Law of Large Numbers* (LLN). The LLN characterizes the behavior of sample averages in relation to sample size. Specifically, the LLN says that a sample average can be brought as close as we like to the average in the population from which it is drawn (say, the population of American college students) simply by enlarging the sample.

To see the LLN in action, play dice.[6] Specifically, roll a fair die once and save the result. Then roll again and average these two results. Keep on rolling and averaging. The numbers 1 to 6 are equally likely (that's why the die is said to be "fair"), so we can expect to see each value an equal number of times if we play long enough. Since there are six possibilities here, and all are equally likely, the expected outcome is an equally weighted average of each possibility, with weights equal to 1/6:

$$\begin{aligned}&\left(1 \times \tfrac{1}{6}\right) + \left(2 \times \tfrac{1}{6}\right) + \left(3 \times \tfrac{1}{6}\right) + \left(4 \times \tfrac{1}{6}\right) + \left(5 \times \tfrac{1}{6}\right) + \left(6 \times \tfrac{1}{6}\right) \\ &= \frac{1+2+3+4+5+6}{6} = 3.5.\end{aligned}$$

This average value of 3.5 is called a *mathematical expectation;* in this case, it's the average value we'd get in infinitely many rolls of a fair die. The expectation concept is important to our work, so we define it formally here.

MATHEMATICAL EXPECTATION The mathematical expectation of a variable, Y_i, written $E[Y_i]$, is the population average of this variable. If Y_i is a variable generated by a random process, such as throwing a die, $E[Y_i]$ is the average in infinitely many repetitions of this process. If Y_i is a variable that comes from a sample survey, $E[Y_i]$ is the average obtained if everyone in the population from which the sample is drawn were to be enumerated.

Rolling a die only a few times, the average toss may be far from the corresponding mathematical expectation. Roll two times, for example, and you might get boxcars or snake eyes (two sixes or two ones). These average to values well away from the expected value of 3.5. But as the number of tosses goes up, the average across tosses reliably tends to 3.5. This is the LLN in action (and it's how casinos make a profit: in most gambling games, you can't beat the house in the long run, because the expected payout for players is negative). More remarkably, it needn't take too many rolls or too large a sample for a sample average to approach the expected value. The chapter appendix addresses the question of how the number of rolls or the size of a sample survey determines statistical accuracy.

In randomized trials, experimental samples are created by sampling from a population we'd like to study rather than by repeating a game, but the LLN works just the same. When sampled subjects are randomly divided (as if by a coin toss) into treatment and control groups, they come from the same underlying population. The LLN therefore promises that those in randomly assigned treatment and control samples will be similar if the samples are large enough. For example, we expect to see similar proportions of men and women in randomly assigned treatment and control groups. Random assignment also produces groups of about the same age and with similar schooling levels. In fact, randomly assigned groups should be similar in every way, including in ways that we cannot easily measure or

observe. This is the root of random assignment's awesome power to eliminate selection bias.

The power of random assignment can be described precisely using the following definition, which is closely related to the definition of mathematical expectation.

CONDITIONAL EXPECTATION The conditional expectation of a variable, Y_i, given a dummy variable, $D_i = 1$, is written $E[Y_i|D_i = 1]$. This is the average of Y_i in the population that has D_i equal to 1. Likewise, the conditional expectation of a variable, Y_i, given $D_i = 0$, written $E[Y_i|D_i = 0]$, is the average of Y_i in the population that has D_i equal to 0. If Y_i and D_i are variables generated by a random process, such as throwing a die under different circumstances, $E[Y_i|D_i = d]$ is the average of infinitely many repetitions of this process while holding the circumstances indicated by D_i fixed at d. If Y_i and D_i come from a sample survey, $E[Y_i|D_i = d]$ is the average computed when everyone in the population who has $D_i = d$ is sampled.

Because randomly assigned treatment and control groups come from the same underlying population, they are the same in every way, including their expected Y_{0i}. In other words, the conditional expectations, $E[Y_{0i}|D_i = 1]$ and $E[Y_{0i}|D_i = 0]$, are the same. This in turn means that:

RANDOM ASSIGNMENT ELIMINATES SELECTION BIAS When D_i is randomly assigned, $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$, and the difference in expectations by treatment status captures the causal effect of treatment:

$$\begin{aligned}
&E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
&= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
&= E[Y_{0i} + \kappa|D_i = 1] - E[Y_{0i}|D_i = 0] \\
&= \kappa + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
&= \kappa.
\end{aligned}$$

Provided the sample at hand is large enough for the LLN to work its magic (so we can replace the conditional averages in equation (1.4)

with conditional expectations), selection bias disappears in a randomized experiment. Random assignment works not by eliminating individual differences but rather by ensuring that the mix of individuals being compared is the same. Think of this as comparing barrels that include equal proportions of apples and oranges. As we explain in the chapters that follow, randomization isn't the only way to generate such *ceteris paribus* comparisons, but most masters believe it's the best.

When analyzing data from a randomized trial or any other research design, masters almost always begin with a check on whether treatment and control groups indeed look similar. This process, called *checking for balance,* amounts to a comparison of sample averages as in panel B of Table 1.1. The average characteristics in panel B appear dissimilar or unbalanced, underlining the fact that the data in this table don't come from anything like an experiment. It's worth checking for balance in this manner any time you find yourself estimating causal effects.

Random assignment of health insurance seems like a fanciful proposition. Yet health insurance coverage has twice been randomly assigned to large representative samples of Americans. The RAND Health Insurance Experiment (HIE), which ran from 1974 to 1982, was one of the most influential social experiments in research history. The HIE enrolled 3,958 people aged 14 to 61 from six areas of the country. The HIE sample excluded Medicare participants and most Medicaid and military health insurance subscribers. HIE participants were randomly assigned to one of 14 insurance plans. Participants did not have to pay insurance premiums, but the plans had a variety of provisions related to cost sharing, leading to large differences in the amount of insurance they offered.

The most generous HIE plan offered comprehensive care for free. At the other end of the insurance spectrum, three "catastrophic coverage" plans required families to pay 95% of their health-care costs, though these costs were capped as a proportion of income (or capped at $1,000 per family, if that was lower). The catastrophic plans approximate a no-insurance condition. A second insurance

scheme (the “individual deductible” plan) also required families to pay 95% of outpatient charges, but only up to $150 per person or $450 per family. A group of nine other plans had a variety of coinsurance provisions, requiring participants to cover anywhere from 25% to 50% of charges, but always capped at a proportion of income or $1,000, whichever was lower. Participating families enrolled in the experimental plans for 3 or 5 years and agreed to give up any earlier insurance coverage in return for a fixed monthly payment unrelated to their use of medical care.[7]

The HIE was motivated primarily by an interest in what economists call the price elasticity of demand for health care. Specifically, the RAND investigators wanted to know whether and by how much health-care use falls when the price of health care goes up. Families in the free care plan faced a price of zero, while coinsurance plans cut prices to 25% or 50% of costs incurred, and families in the catastrophic coverage and deductible plans paid something close to the sticker price for care, at least until they hit the spending cap. But the investigators also wanted to know whether more comprehensive and more generous health insurance coverage indeed leads to better health. The answer to the first question was a clear “yes”: health-care consumption is highly responsive to the price of care. The answer to the second question is murkier.

Randomized Results

Randomized field experiments are more elaborate than a coin toss, sometimes regrettably so. The HIE was complicated by having many small treatment groups, spread over more than a dozen insurance plans. The treatment groups associated with each plan are mostly too small for comparisons between them to be statistically meaningful. Most analyses of the HIE data therefore start by grouping subjects who were assigned to similar HIE plans together. We do that here as well.[8]

A natural grouping scheme combines plans by the amount of cost sharing they require. The three catastrophic coverage plans, with subscribers shouldering almost all of their medical expenses up to a fairly high cap, approximate a no-insurance state. The individual

deductible plan provided more coverage, but only by reducing the cap on total expenses that plan participants were required to shoulder. The nine coinsurance plans provided more substantial coverage by splitting subscribers' health-care costs with the insurer, starting with the first dollar of costs incurred. Finally, the free plan constituted a radical intervention that might be expected to generate the largest increase in health-care usage and, perhaps, health. This categorization leads us to four groups of plans: catastrophic, deductible, coinsurance, and free, instead of the 14 original plans. The catastrophic plans provide the (approximate) no-insurance control, while the deductible, coinsurance, and free plans are characterized by increasing levels of coverage.

As with nonexperimental comparisons, a first step in our experimental analysis is to check for balance. Do subjects randomly assigned to treatment and control groups—in this case, to health insurance schemes ranging from little to complete coverage—indeed look similar? We gauge this by comparing demographic characteristics and health data collected before the experiment began. Because demographic characteristics are unchanging, while the health variables in question were measured before random assignment, we expect to see only small differences in these variables across the groups assigned to different plans.

In contrast with our comparison of NHIS respondents' characteristics by insurance status in Table 1.1, a comparison of characteristics across randomly assigned treatment groups in the RAND experiment shows the people assigned to different HIE plans to be similar. This can be seen in panel A of Table 1.3. Column (1) in this table reports averages for the catastrophic plan group, while the remaining columns compare the groups assigned more generous insurance coverage with the catastrophic control group. As a summary measure, column (5) compares a sample combining subjects in the deductible, coinsurance, and free plans with subjects in the catastrophic plans. Individuals assigned to the plans with more generous coverage are a little less likely to be female and a little less educated than those in the catastrophic plans. We also see some variation in income, but differences between plan groups are mostly

small and are as likely to go one way as another. This pattern contrasts with the large and systematic demographic differences between insured and uninsured people seen in the NHIS data summarized in Table 1.1.

The small differences across groups seen in panel A of Table 1.3 seem likely to reflect chance variation that emerges naturally as part of the sampling process. In any statistical sample, chance differences arise because we're looking at one of many possible draws from the underlying population from which we've sampled. A new sample of similar size from the same population can be expected to produce comparisons that are similar—though not identical—to those in the table. The question of how much variation we should expect from one sample to another is addressed by the tools of statistical inference.

TABLE 1.3

Demographic characteristics and baseline health in the RAND HIE

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
A. Demographic characteristics					
Female	.560	−.023	−.025	−.038	−.030
		(.016)	(.015)	(.015)	(.013)
Nonwhite	.172	−.019	−.027	−.028	−.025
		(.027)	(.025)	(.025)	(.022)
Age	32.4	.56	.97	.43	.64
	[12.9]	(.68)	(.65)	(.61)	(.54)
Education	12.1	−.16	−.06	−.26	−.17
	[2.9]	(.19)	(.19)	(.18)	(.16)
Family income	31,603	−2,104	970	−976	−654
	[18,148]	(1,384)	(1,389)	(1,345)	(1,181)
Hospitalized last year	.115	.004	−.002	.001	.001
		(.016)	(.015)	(.015)	(.013)
B. Baseline health variables					
General health index	70.9	−1.44	.21	−1.31	−.93
	[14.9]	(.95)	(.92)	(.87)	(.77)
Cholesterol (mg/dl)	207	−1.42	−1.93	−5.25	−3.19
	[40]	(2.99)	(2.76)	(2.70)	(2.29)
Systolic blood pressure (mm Hg)	122	2.32	.91	1.12	1.39
	[17]	(1.15)	(1.08)	(1.01)	(.90)
Mental health index	73.8	−.12	1.19	.89	.71
	[14.3]	(.82)	(.81)	(.77)	(.68)
Number enrolled	759	881	1,022	1,295	3,198

Notes: This table describes the demographic characteristics and baseline health of subjects in the RAND Health Insurance Experiment (HIE). Column (1) shows the average for the group assigned catastrophic coverage. Columns (2)–(5) compare averages in the deductible, cost-sharing, free care, and any insurance groups with the average in column (1). Standard errors are reported in parentheses in columns (2)–(5); standard deviations are reported in brackets in column (1).

The appendix to this chapter briefly explains how to quantify sampling variation with formal statistical tests. Such tests amount to the juxtaposition of differences in sample averages with their *standard errors,* the numbers in parentheses reported below the differences in averages listed in columns (2)–(5) of Table 1.3. The standard error of a difference in averages is a measure of its statistical precision: when a difference in sample averages is smaller than about two standard errors, the difference is typically judged to be a chance finding compatible with the hypothesis that the populations from which these

samples were drawn are, in fact, the same.

Differences that are larger than about two standard errors are said to be *statistically significant:* in such cases, it is highly unlikely (though not impossible) that these differences arose purely by chance. Differences that are not statistically significant are probably due to the vagaries of the sampling process. The notion of statistical significance helps us interpret comparisons like those in Table 1.3. Not only are the differences in this table mostly small, only two (for proportion female in columns (4) and (5)) are more than twice as large as the associated standard errors. In tables with many comparisons, the presence of a few isolated statistically significant differences is usually also attributable to chance. We also take comfort from the fact that the standard errors in this table are not very big, indicating differences across groups are measured reasonably precisely.

Panel B of Table 1.3 complements the contrasts in panel A with evidence for reasonably good balance in *pre-treatment outcomes* across treatment groups. This panel shows no statistically significant differences in a pre-treatment index of general health. Likewise, pre-treatment cholesterol, blood pressure, and mental health appear largely unrelated to treatment assignment, with only a couple of contrasts close to statistical significance. In addition, although lower cholesterol in the free group suggests somewhat better health than in the catastrophic group, differences in the general health index between these two groups go the other way (since lower index values indicate worse health). Lack of a consistent pattern reinforces the notion that these gaps are due to chance.

The first important finding to emerge from the HIE was that subjects assigned to more generous insurance plans used substantially more health care. This finding, which vindicates economists' view that demand for a good should go up when it gets cheaper, can be seen in panel A of Table 1.4.[9] As might be expected, hospital inpatient admissions were less sensitive to price than was outpatient care, probably because admissions decisions are usually made by doctors. On the other hand, assignment to the free care plan raised outpatient spending by two-thirds (169/248) relative to spending by those in

catastrophic plans, while total medical expenses increased by 45%. These large gaps are economically important as well as statistically significant.

Subjects who didn't have to worry about the cost of health care clearly consumed quite a bit more of it. Did this extra care and expense make them healthier? Panel B in Table 1.4, which compares health indicators across HIE treatment groups, suggests not. Cholesterol levels, blood pressure, and summary indices of overall health and mental health are remarkably similar across groups (these outcomes were mostly measured 3 or 5 years after random assignment). Formal statistical tests show no statistically significant differences, as can be seen in the group-specific contrasts (reported in columns (2)–(4)) and in the differences in health between those in a catastrophic plan and everyone in the more generous insurance groups (reported in column (5)).

These HIE findings convinced many economists that generous health insurance can have unintended and undesirable consequences, increasing health-care usage and costs, without generating a dividend in the form of better health.[10]

TABLE 1.4

Health expenditure and health outcomes in the RAND HIE

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
		A. Health-care use			
Face-to-face visits	2.78	.19	.48	1.66	.90
	[5.50]	(.25)	(.24)	(.25)	(.20)
Outpatient expenses	248	42	60	169	101
	[488]	(21)	(21)	(20)	(17)
Hospital admissions	.099	.016	.002	.029	.017
	[.379]	(.011)	(.011)	(.010)	(.009)
Inpatient expenses	388	72	93	116	97
	[2,308]	(69)	(73)	(60)	(53)
Total expenses	636	114	152	285	198
	[2,535]	(79)	(85)	(72)	(63)
		B. Health outcomes			
General health index	68.5	−.87	.61	−.78	−.36
	[15.9]	(.96)	(.90)	(.87)	(.77)
Cholesterol (mg/dl)	203	.69	−2.31	−1.83	−1.32
	[42]	(2.57)	(2.47)	(2.39)	(2.08)
Systolic blood pressure (mm Hg)	122	1.17	−1.39	−.52	−.36
	[19]	(1.06)	(.99)	(.93)	(.85)
Mental health index	75.5	.45	1.07	.43	.64
	[14.8]	(.91)	(.87)	(.83)	(.75)
Number enrolled	759	881	1,022	1,295	3,198

Notes: This table reports means and treatment effects for health expenditure and health outcomes in the RAND Health Insurance Experiment (HIE). Column (1) shows the average for the group assigned catastrophic coverage. Columns (2)–(5) compare averages in the deductible, cost-sharing, free care, and any insurance groups with the average in column (1). Standard errors are reported in parentheses in columns (2)–(5); standard deviations are reported in brackets in column (1).

1.2 The Oregon Trail

MASTER KAN: Truth is hard to understand.

KWAI CHANG CAINE: It is a fact, it is not the truth. Truth is often hidden, like a shadow in darkness.

Kung Fu, Season 1, Episode 14

The HIE was an ambitious attempt to assess the impact of health insurance on health-care costs and health. And yet, as far as the contemporary debate over health insurance goes, the HIE might have missed the mark. For one thing, each HIE treatment group had at least catastrophic coverage, so financial liability for health-care costs was

limited under every treatment. More importantly, today's uninsured Americans differ considerably from the HIE population: most of the uninsured are younger, less educated, poorer, and less likely to be working. The value of extra health care in such a group might be very different than for the middle class families that participated in the HIE.

One of the most controversial ideas in the contemporary health policy arena is the expansion of Medicaid to cover the currently uninsured (interestingly, on the eve of the RAND experiment, talk was of expanding Medicare, the public insurance program for America's elderly). Medicaid now covers families on welfare, some of the disabled, other poor children, and poor pregnant women. Suppose we were to expand Medicaid to cover those who don't qualify under current rules. How would such an expansion affect health-care spending? Would it shift treatment from costly and crowded emergency departments to possibly more effective primary care? Would Medicaid expansion improve health?

Many American states have begun to "experiment" with Medicaid expansion in the sense that they've agreed to broaden eligibility, with the federal government footing most of the bill. Alas, these aren't real experiments, since everyone who is eligible for expanded Medicaid coverage gets it. The most convincing way to learn about the consequences of Medicaid expansion is to randomly offer Medicaid coverage to people in currently ineligible groups. Random assignment of Medicaid seems too much to hope for. Yet, in an awesome social experiment, the state of Oregon recently offered Medicaid to thousands of randomly chosen people in a publicly announced health insurance lottery.

We can think of Oregon's health insurance lottery as randomly selecting winners and losers from a pool of registrants, though coverage was not automatic, even for lottery winners. Winners won the opportunity to apply for the state-run Oregon Health Plan (OHP), the Oregon version of Medicaid. The state then reviewed these applications, awarding coverage to Oregon residents who were U.S. citizens or legal immigrants aged 19–64, not otherwise eligible for

Medicaid, uninsured for at least 6 months, with income below the federal poverty level, and few financial assets. To initiate coverage, lottery winners had to document their poverty status and submit the required paperwork within 45 days.

The rationale for the 2008 OHP lottery was fairness and not research, but it's no less awesome for that. The Oregon health insurance lottery provides some of the best evidence we can hope to find on the costs and benefits of insurance coverage for the currently uninsured, a fact that motivated research on OHP by MIT master Amy Finkelstein and her coauthors.[11]

Roughly 75,000 lottery applicants registered for expanded coverage through the OHP. Of these, almost 30,000 were randomly selected and invited to apply for OHP; these winners constitute the OHP treatment group. The other 45,000 constitute the OHP control sample.

The first question that arises in this context is whether OHP lottery winners were more likely to end up insured as a result of winning. This question is motivated by the fact that some applicants qualified for regular Medicaid even without the lottery. Panel A of Table 1.5 shows that about 14% of controls (lottery losers) were covered by Medicaid in the year following the first OHP lottery. At the same time, the second column, which reports differences between the treatment and control groups, shows that the probability of Medicaid coverage increased by 26 percentage points for lottery winners. Column (4) shows a similar increase for the subsample living in and around Portland, Oregon's largest city. The upshot is that OHP lottery winners were insured at much higher rates than were lottery losers, a difference that might have affected their use of health care and their health.[12]

The OHP treatment group (that is, lottery winners) used more health-care services than they otherwise would have. This can also be seen in Table 1.5, which shows estimates of changes in service use in the rows below the estimate of the OHP effect on Medicaid coverage. The hospitalization rate increased by about half a percentage point, a modest though statistically significant effect. Emergency department visits, outpatient visits, and prescription drug use all increased

markedly. The fact that the number of emergency department visits rose about 10%, a precisely estimated effect (the standard error associated with this estimate, reported in column (4), is .029), is especially noteworthy. Many policymakers hoped and expected health insurance to shift formerly uninsured patients away from hospital emergency departments toward less costly sources of care.

TABLE 1.5

OHP effects on insurance coverage and health-care use

	Oregon		Portland area	
Outcome	Control mean (1)	Treatment effect (2)	Control mean (3)	Treatment effect (4)
A. Administrative data				
Ever on Medicaid	.141	.256 (.004)	.151	.247 (.006)
Any hospital admissions	.067	.005 (.002)		
Any emergency department visit			.345	.017 (.006)
Number of emergency department visits			1.02	.101 (.029)
Sample size	74,922		24,646	
B. Survey data				
Outpatient visits (in the past 6 months)	1.91	.314 (.054)		
Any prescriptions?	.637	.025 (.008)		
Sample size	23,741			

Notes: This table reports estimates of the effect of winning the Oregon Health Plan (OHP) lottery on insurance coverage and use of health care. Odd-numbered columns show control group averages. Even-numbered columns report the regression coefficient on a dummy for lottery winners. Standard errors are reported in parentheses.

Finally, the proof of the health insurance pudding appears in Table 1.6: lottery winners in the statewide sample report a modest improvement in the probability they assess their health as being good or better (an effect of .039, which can be compared with a control mean of .55; the Health is Good variable is a dummy). Results from

in-person interviews conducted in Portland suggest these gains stem more from improved mental rather than physical health, as can be seen in the second and third rows in column (4) (the health variables in the Portland sample are indices ranging from 0 to 100). As in the RAND experiment, results from Portland suggest physical health indicators like cholesterol and blood pressure were largely unchanged by increased access to OHP insurance.

TABLE 1.6

OHP effects on health indicators and financial health

	Oregon		Portland area	
Outcome	Control mean (1)	Treatment effect (2)	Control mean (3)	Treatment effect (4)
A. Health indicators				
Health is good	.548	.039 (.008)		
Physical health index			45.5	.29 (.21)
Mental health index			44.4	.47 (.24)
Cholesterol			204	.53 (.69)
Systolic blood pressure (mm Hg)			119	−.13 (.30)
B. Financial health				
Medical expenditures >30% of income			.055	−.011 (.005)
Any medical debt?			.568	−.032 (.010)
Sample size	23,741		12,229	

Notes: This table reports estimates of the effect of winning the Oregon Health Plan (OHP) lottery on health indicators and financial health. Odd-numbered columns show control group averages. Even-numbered columns report the regression coefficient on a dummy for lottery winners. Standard errors are reported in parentheses.

The weak health effects of the OHP lottery disappointed policymakers who looked to publicly provided insurance to generate a health dividend for low-income Americans. The fact that health insurance increased rather than decreased expensive emergency

department use is especially frustrating. At the same time, panel B of Table 1.6 reveals that health insurance provided the sort of financial safety net for which it was designed. Specifically, households winning the lottery were less likely to have incurred large medical expenses or to have accumulated debt generated by the need to pay for health care. It may be this improvement in financial health that accounts for improved mental health in the treatment group.

It also bears emphasizing that the financial and health effects seen in Table 1.6 most likely come from the 25% of the sample who obtained insurance as a result of the lottery. Adjusting for the fact that insurance status was unchanged for many winners shows that gains in financial security and mental health for the one-quarter of applicants who were insured as a result of the lottery were considerably larger than simple comparisons of winners and losers would suggest. Chapter 3, on instrumental variables methods, details the nature of such adjustments. As you'll soon see, the appropriate adjustment here amounts to the division of win/loss differences in outcomes by win/loss differences in the probability of insurance. This implies that the effect of being insured is as much as four times larger than the effect of winning the OHP lottery (statistical significance is unchanged by this adjustment).

The RAND and Oregon findings are remarkably similar. Two ambitious experiments targeting substantially different populations show that the use of health-care services increases sharply in response to insurance coverage, while neither experiment reveals much of an insurance effect on physical health. In 2008, OHP lottery winners enjoyed small but noticeable improvements in mental health. Importantly, and not coincidentally, OHP also succeeded in insulating many lottery winners from the financial consequences of poor health, just as a good insurance policy should. At the same time, these studies suggest that subsidized public health insurance should not be expected to yield a dramatic health dividend.

MASTER JOSHWAY: In a nutshell, please, Grasshopper.

GRASSHOPPER: Causal inference compares potential outcomes,

descriptions of the world when alternative roads are taken.

MASTER JOSHWAY: Do we compare those who took one road with those who took another?

GRASSHOPPER: Such comparisons are often contaminated by selection bias, that is, differences between treated and control subjects that exist even in the absence of a treatment effect.

MASTER JOSHWAY: Can selection bias be eliminated?

GRASSHOPPER: Random assignment to treatment and control conditions eliminates selection bias. Yet even in randomized trials, we check for balance.

MASTER JOSHWAY: Is there a single causal truth, which all randomized investigations are sure to reveal?

GRASSHOPPER: I see now that there can be many truths, Master, some compatible, some in contradiction. We therefore take special note when findings from two or more experiments are similar.

Masters of 'Metrics: From Daniel to R. A. Fisher

The value of a control group was revealed in the Old Testament. The Book of Daniel recounts how Babylonian King Nebuchadnezzar decided to groom Daniel and other Israelite captives for his royal service. As slavery goes, this wasn't a bad gig, since the king ordered his captives be fed "food and wine from the king's table." Daniel was uneasy about the rich diet, however, preferring modest vegetarian fare. The king's chamberlains initially refused Daniel's special meals request, fearing that his diet would prove inadequate for one called on to serve the king. Daniel, not without chutzpah, proposed a controlled experiment: "Test your servants for ten days. Give us nothing but vegetables to eat and water to drink. Then compare our appearance with that of the young men who eat the royal food, and treat your servants in accordance with what you see" (Daniel 1, 12–13). The Bible recounts how this experiment supported Daniel's conjecture regarding the relative healthfulness of a vegetarian diet, though as far as we know Daniel himself didn't get an academic paper out of it.

Nutrition is a recurring theme in the quest for balance. Scurvy, a

debilitating disease caused by vitamin C deficiency, was the scourge of the British Navy. In 1742, James Lind, a surgeon on HMS *Salisbury,* experimented with a cure for scurvy. Lind chose 12 seamen with scurvy and started them on an identical diet. He then formed six pairs and treated each of the pairs with a different supplement to their daily food ration. One of the additions was an extra two oranges and one lemon (Lind believed an acidic diet might cure scurvy). Though Lind did not use random assignment, and his sample was small by our standards, he was a pioneer in that he chose his 12 study members so they were "as similar as I could have them." The citrus eaters—Britain's first limeys—were quickly and incontrovertibly cured, a life-changing empirical finding that emerged from Lind's data even though his theory was wrong.[13]

Almost 150 years passed between Lind and the first recorded use of experimental random assignment. This was by Charles Peirce, an American philosopher and scientist, who experimented with subjects' ability to detect small differences in weight. In a less-than-fascinating but methodologically significant 1885 publication, Peirce and his student Joseph Jastrow explained how they varied experimental conditions according to draws from a pile of playing cards.[14]

The idea of a randomized controlled trial emerged in earnest only at the beginning of the twentieth century, in the work of statistician and geneticist Sir Ronald Aylmer Fisher, who analyzed data from agricultural experiments. Experimental random assignment features in

Fisher's 1925 *Statistical Methods for Research Workers* and is detailed in his landmark *The Design of Experiments,* published in 1935.[15]

Fisher had many fantastically good ideas and a few bad ones. In addition to explaining the value of random assignment, he invented the statistical method of maximum likelihood. Along with 'metrics master Sewall Wright (and J.B.S. Haldane), he launched the field of theoretical population genetics. But he was also a committed eugenicist and a proponent of forced sterilization (as was regression master Sir Francis Galton, who coined the term "eugenics"). Fisher, a lifelong pipe smoker, was also on the wrong side of the debate over smoking and health, due in part to his strongly held belief that smoking and lung cancer share a common genetic origin. The negative effect of smoking on health now seems well established, though Fisher was right to worry about selection bias in health research. Many lifestyle choices, such as low-fat diets and vitamins, have been shown to be unrelated to health outcomes when evaluated with random assignment.

Appendix: Mastering Inference

YOUNG CAINE: I am puzzled.

MASTER PO: That is the beginning of wisdom.

Kung Fu, Season 2, Episode 25

This is the first of a number of appendices that fill in key econometric and statistical details. You can spend your life studying statistical inference; many masters do. Here we offer a brief sketch of essential ideas and basic statistical tools, enough to understand tables like those in this chapter.

The HIE is based on a sample of participants drawn (more or less) at random from the population eligible for the experiment. Drawing another sample from the same population, we'd get somewhat different results, but the general picture should be similar if the sample is large enough for the LLN to kick in. How can we decide whether statistical results constitute strong evidence or merely a lucky draw, unlikely to be replicated in repeated samples? How much sampling variance should we expect? The tools of formal statistical