Chapter Two

The Trouble with Scientists, Part 1

If we knew what we were doing, it wouldn’t be called research, would it?
— Albert Einstein

When you can measure what you are speaking about, and express it in numbers, you know something about it.
— Lord Kelvin

Philip “P.J.” Devereaux, a cardiologist and biostatistics researcher at McMaster University in Ontario, remembers all too well the bolt of excitement that ran through the field of cardiology back in the early 1980s. Cardiologists had already come to recognize that irregular heartbeats were an ominous sign, observing that those who had them within the first twelve days after a heart attack were far more likely to die than those who didn’t. Then antiarrhythmic drugs burst onto the scene, and cardiologists seized on them. They gave the medication to heart-attack patients, closely monitored their heart rhythms, and were thrilled to see the heart rhythms smooth out. The fast administration of an antiarrhythmic drug quickly became the standard treatment for heart-attack patients, and remained so through the early 1990s, when the results of a new, large study of these drugs came out. “The trial didn’t just show that the drugs weren’t saving lives,
it showed they actually were killing people,” says Devereaux. Yes, the patients’ hearts were beating more regularly on the drug, but the patients were on average three times more likely to die. In fact, notes Devereaux, the drugs killed more Americans than the Vietnam War did—roughly an average of forty thousand a year died from the drugs in the United States alone. Where had cardiologists gone wrong?

It would be unusual to make it through a Good Morning America broadcast or two, or an issue of the Boston Globe, or a day’s worth of Yahoo!’s home page, or the like, without being exposed to the considered conclusions of at least one highly degreed scientist or other academic researcher. Not only are the latest medical studies fairly sprayed at us from all directions, but in the course of a week of run-of-the-mill browsing of print, broadcast, and Internet news and other information sources, you’re likely to encounter a parade of confident PhDs that might include economists (How long will the recession last?), psychologists (Is your teenager depressed?), sociologists (Is our cell phone obsession ruining society?), zoologists (Can the polar bear be saved?), and more, all obligingly helping to translate their or their field’s latest research findings into news you can use. What’s more, scientific research studies often underlie the advice we hear from less formal experts, be it the nutrition editor at a women’s health magazine, the head of an industry or consumer group, the personal trainer to the stars, the hotshot portfolio manager, the retired general, the author of a book on relationships, the former U.S. vice president, or the celebrated ex-CEO of a successful company. There’s a reason that the phrases “experts say,” “studies show,” and “according to the latest research” all sound like well-worn clichés to us.

Given that much of what we are told about the world is built one way or another on published research, we seem to have a big problem. John Ioannidis’s work suggests that wrongness is the rule rather than the exception, and especially in medical research, which has an outsize impact on our lives, gets more than its share of attention in the media, and attracts a stupendous depth of talent and funding—about $95 billion a year, or 6 percent of all money spent on health care, according to a 2005 study. While experts who study research may have become comfortable with the notion that most findings are wrong, it might seem a little hard to swallow for the rest of us. It would help if we had some sense of why so much research might be turning out to be wrong.

Scientists and science journalists often dismiss wayward scientific findings as the product of weaker types of studies. But the fact is, it’s absolutely typical for studies that have all the markings of high credibility to contradict one another or simply get it wrong. And here’s a simple reason: researchers routinely rely on flawed evidence in coming to their conclusions and in working to convince us that those conclusions are right. To put it another way, scientists are often deceptively sloppy in making and analyzing measurements. And that’s in spite of the fact that good measurement is at the heart of what separates the respected, high-level expert from the opinionated diletante, pop guru, manipulative charlatan, blathering pundit, or junk scientist. Let’s go on a tour of some of the ways impressive-seeming measurements can and often do go wrong for our most-trusted experts.

Measuring What Doesn’t Matter

Can vitamin D supplements help fend off cancer?
No, said a 1999 study.2
Yes, said a study in 2006— it cuts risk up to 50 percent.3
Yes, said a study in 2007— it cuts risk up to 77 percent.4
No, said a 2008 study.5
You'd almost have to laugh at these sorts of seesaw, yes-it-is/no-it-isn't findings, if they weren't addressing potentially life-and-death questions. Nearly half of us are going to get cancer at some point in our lives, more than a quarter of us will die from it, and those of us who don't get it will still be deeply affected by it, probably by living in fear of it, almost certainly by losing someone to it. Can't we get a little straight advice here?

An old joke: A police officer finds a drunk man late at night crawling on his hands and knees on a sidewalk under a streetlight. Questioned, the drunk man tells her he's looking for his wallet. When the officer asks if he's sure that he dropped the wallet here, the man replies that he actually dropped it across the street. "Then why are you looking over here?" asks the befuddled officer. "Because the light's better here," explains the drunk man.

It's easy to appreciate the foolishness of choosing to search for something in a way that's relatively convenient rather than in a way that's more likely to be fruitful. But experts do it all the time when it comes to searching for the truth — in fact, in some fields, as we'll see, they almost always end up looking under the streetlight. In the case of antiarrhythmic drugs for heart-attack victims, cardiologists knew that what they really wanted to measure was survival. But measuring survival takes long, relatively complex studies. Having observed an apparent link between irregular heartbeats and death, they found it reasonable to jump on irregular beats as a relevant measurement, under the assumption that as the frequency of irregular beats came down, so would death rates. The irregular beats were measured, and doctors saw they were successfully suppressed by the drugs — mission accomplished! But the notion that suppressing irregular heartbeats would keep heart-attack patients alive was simply wrong, and in a particularly deadly way. Or take the 1999 study mentioned previously, which concluded that vitamin D had no effect on the risk of breast cancer. Instead of directly measuring vitamin D, the study, it turned out, had estimated vitamin D levels in the body by relying on what patients reported for diets and on the estimated amounts of sunlight each absorbed as based on geographic location (in that sunlight spurs the body to manufacture vitamin D).

These sorts of indirect measurements are sometimes called "surrogate" or "proxy" measurements, or "markers." Such a measurement is made to stand in for what you really want to measure, typically because it's more accessible in some way — it's more convenient to obtain or can be achieved more cheaply or quickly. Experts often base the bulk of their conclusions on surrogate measurements, and they're especially ubiquitous in medical studies. Instead of having to wait to directly measure cancer survival, researchers have long considered tumor shrinkage to mean that a cancer treatment is effective; lowered blood sugar levels have been considered a sign that the slowly progressing ravages of diabetes are under control; brain scans that show good blood volume are sometimes taken as evidence of the halting of the gradual loss of cognitive function in Alzheimer's; and control of cholesterol levels has stood in for pushing back heart disease and stroke vulnerability. Ioannidis notes, for example, that 21 different studies of asthma he looked at measured a combined 487 different factors in patients in struggling to determine what constitutes genuine improvement; every researcher seemed to have her own idea on the question.