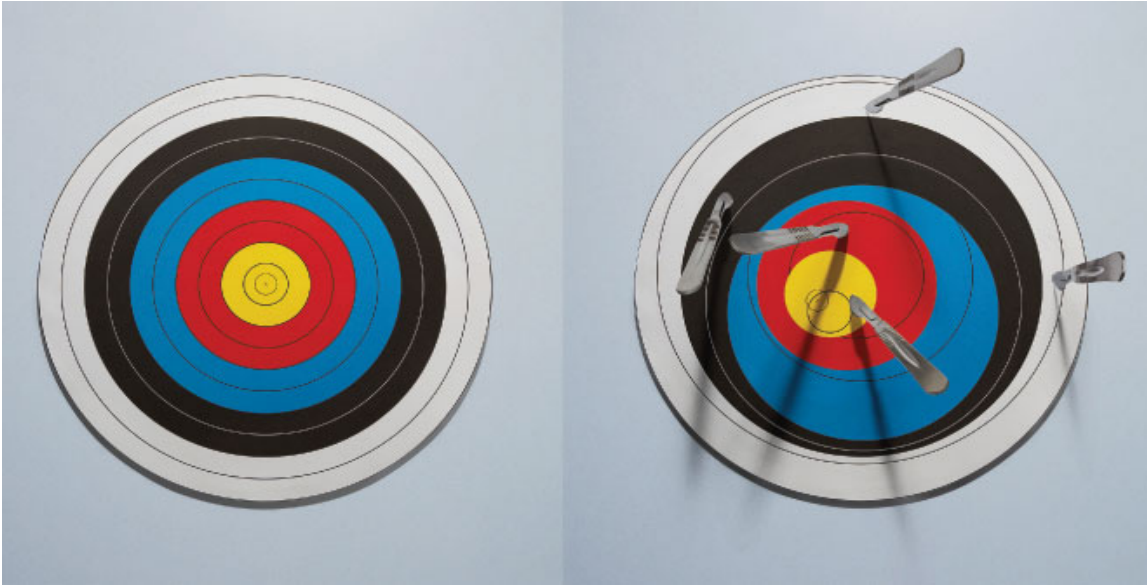


# Trials and Errors: Why Science Is Failing Us

- By Jonah Lehrer
- [Wired January 2012](#) , pp. 102-109.



Every year, nearly \$100 billion is invested in biomedical research in the US, all of it aimed at teasing apart the invisible bits of the body.

Photo: Mauricio Alejo

**On November 30, 2006**, executives at Pfizer—the largest pharmaceutical company in the world—held a meeting with investors at the firm’s research center in Groton, Connecticut. Jeff Kindler, then CEO of Pfizer, began the presentation with an upbeat assessment of the company’s efforts to bring new drugs to market. He cited “exciting approaches” to the treatment of Alzheimer’s disease, fibromyalgia, and arthritis. But that news was just a warm-up. Kindler was most excited about a new drug called torcetrapib, which had recently entered Phase III clinical trials, the last step before filing for FDA approval. He confidently declared that torcetrapib would be “one of the most important compounds of our generation.”

Kindler’s enthusiasm was understandable: The potential market for the drug was enormous. Like Pfizer’s blockbuster medication, Lipitor—the most widely prescribed branded pharmaceutical in America—torcetrapib was designed to tweak the cholesterol pathway. Although cholesterol is an essential component of cellular membranes, high levels of the compound have been consistently associated with heart disease. The accumulation of the pale yellow substance in arterial walls leads to inflammation.

Clusters of white blood cells then gather around these “plaques,” which leads to even more inflammation. The end result is a blood vessel clogged with clumps of fat.

Lipitor works by inhibiting an enzyme that plays a key role in the production of cholesterol in the liver. In particular, the drug lowers the level of low-density lipoprotein (LDL), or so-called bad cholesterol. In recent years, however, scientists have begun to focus on a separate part of the cholesterol pathway, the one that produces high-density lipoproteins. One function of HDL is to transport excess LDL back to the liver, where it is broken down. In essence, HDL is a janitor of fat, cleaning up the greasy mess of the modern diet, which is why it’s often referred to as “good cholesterol.”

And this returns us to torcetrapib. It was designed to block a protein that converts HDL cholesterol into its more sinister sibling, LDL. In theory, this would cure our cholesterol problems, creating a surplus of the good stuff and a shortage of the bad. In his presentation, Kindler noted that torcetrapib had the potential to “redefine cardiovascular treatment.”

There was a vast amount of research behind Kindler’s bold proclamations. The cholesterol pathway is one of the best-understood biological feedback systems in the human body. Since 1913, when Russian pathologist Nikolai Anichkov first experimentally linked cholesterol to the buildup of plaque in arteries, scientists have mapped out the metabolism and transport of these compounds in exquisite detail. They’ve documented the interactions of nearly every molecule, the way hydroxymethylglutaryl-coenzyme A reductase catalyzes the production of mevalonate, which gets phosphorylated and condensed before undergoing a sequence of electron shifts until it becomes lanosterol and then, after another 19 chemical reactions, finally morphs into cholesterol. Furthermore, torcetrapib had already undergone a small clinical trial, which showed that the drug could increase HDL and decrease LDL. Kindler told his investors that, by the second half of 2007, Pfizer would begin applying for approval from the FDA. The success of the drug seemed like a sure thing.

And then, just two days later, on December 2, 2006, Pfizer issued a stunning announcement: The torcetrapib Phase III clinical trial was being terminated. Although the compound was supposed to prevent heart disease, it was actually triggering higher rates of chest pain and heart failure and a 60 percent increase in overall mortality. The drug appeared to be killing people.

That week, Pfizer’s value plummeted by \$21 billion.

**The story of torcetrapib** is a tale of mistaken causation. Pfizer was operating on the assumption that raising levels of HDL cholesterol and lowering LDL would lead to a predictable outcome: Improved cardiovascular health. Less arterial plaque. Cleaner pipes. But that didn’t happen.

Such failures occur all the time in the drug industry. (According to one recent analysis, more than 40 percent of drugs fail Phase III clinical trials.) And yet there is something

particularly disturbing about the failure of torcetrapib. After all, a bet on this compound wasn't supposed to be risky. For Pfizer, torcetrapib was the payoff for decades of research. Little wonder that the company was so confident about its clinical trials, which involved a total of 25,000 volunteers. Pfizer invested more than \$1 billion in the development of the drug and \$90 million to expand the factory that would manufacture the compound. Because scientists understood the individual steps of the cholesterol pathway at such a precise level, they assumed they also understood how it worked as a whole.

This assumption—that understanding a system's constituent parts means we also understand the causes within the system—is not limited to the pharmaceutical industry or even to biology. It defines modern science. In general, we believe that the so-called problem of causation can be cured by more information, by our ceaseless accumulation of facts. Scientists refer to this process as reductionism. By breaking down a process, we can see how everything fits together; the complex mystery is distilled into a list of ingredients. And so the question of cholesterol—what is its relationship to heart disease?—becomes a predictable loop of proteins tweaking proteins, acronyms altering one another. Modern medicine is particularly reliant on this approach. Every year, nearly \$100 billion is invested in biomedical research in the US, all of it aimed at teasing apart the invisible bits of the body. We assume that these new details will finally reveal the causes of illness, pinning our maladies on small molecules and errant snippets of DNA. Once we find the cause, of course, we can begin working on a cure.



Photo: Mauricio Alejo

The problem with this assumption, however, is that causes are a strange kind of knowledge. This was first pointed out by David Hume, the 18th-century Scottish philosopher. Hume realized that, although people talk about causes as if they are real facts—tangible things that can be discovered—they're actually not at all factual. Instead, Hume said, every cause is just a slippery story, a catchy conjecture, a “lively conception

produced by habit.” When an apple falls from a tree, the cause is obvious: gravity. Hume’s skeptical insight was that we don’t see gravity—we see only an object tugged toward the earth. We look at X and then at Y, and invent a story about what happened in between. We can measure facts, but a cause is not a fact—it’s a fiction that helps us make sense of facts.

The truth is, our stories about causation are shadowed by all sorts of mental shortcuts. Most of the time, these shortcuts work well enough. They allow us to hit fastballs, discover the law of gravity, and design wondrous technologies. However, when it comes to reasoning about complex systems—say, the human body—these shortcuts go from being slickly efficient to outright misleading.

Consider a set of classic experiments designed by Belgian psychologist Albert Michotte, first conducted in the 1940s. The research featured a series of short films about a blue ball and a red ball. In the first film, the red ball races across the screen, touches the blue ball, and then stops. The blue ball, meanwhile, begins moving in the same basic direction as the red ball. When Michotte asked people to describe the film, they automatically lapsed into the language of causation. The red ball hit the blue ball, which *caused* it to move.

This is known as the launching effect, and it’s a universal property of visual perception. Although there was nothing about causation in the two-second film—it was just a montage of animated images—people couldn’t help but tell a story about what had happened. They translated their perceptions into causal beliefs.

Michotte then began subtly manipulating the films, asking the subjects how the new footage changed their description of events. For instance, when he introduced a one-second pause between the movement of the balls, the impression of causality disappeared. The red ball no longer appeared to trigger the movement of the blue ball. Rather, the two balls were moving for inexplicable reasons.

Michotte would go on to conduct more than 100 of these studies. Sometimes he would have a small blue ball move in front of a big red ball. When he asked subjects what was going on, they insisted that the red ball was “chasing” the blue ball. However, if a big red ball was moving in front of a little blue ball, the opposite occurred: The blue ball was “following” the red ball.

There are two lessons to be learned from these experiments. The first is that our theories about a particular cause and effect are inherently perceptual, infected by all the sensory cheats of vision. (Michotte compared causal beliefs to color perception: We apprehend what we perceive as a cause as automatically as we identify that a ball is red.) While Hume was right that causes are never seen, only inferred, the blunt truth is that we can’t tell the difference. And so we look at moving balls and automatically see causes, a melodrama of taps and collisions, chasing and fleeing.

The second lesson is that causal explanations are oversimplifications. This is what makes them useful—they help us grasp the world at a glance. For instance, after watching the short films, people immediately settled on the most straightforward explanation for the ricocheting objects. Although this account felt true, the brain wasn't seeking the literal truth—it just wanted a plausible story that didn't contradict observation.

This mental approach to causality is often effective, which is why it's so deeply embedded in the brain. However, those same shortcuts get us into serious trouble in the modern world when we use our perceptual habits to explain events that we can't perceive or easily understand. Rather than accept the complexity of a situation—say, that snarl of causal interactions in the cholesterol pathway—we persist in pretending that we're staring at a blue ball and a red ball bouncing off each other. There's a fundamental mismatch between how the world works and how we think about the world.

The good news is that, in the centuries since Hume, scientists have mostly managed to work around this mismatch as they've continued to discover new cause-and-effect relationships at a blistering pace. This success is largely a tribute to the power of statistical correlation, which has allowed researchers to pirouette around the problem of causation. Though scientists constantly remind themselves that mere correlation is *not* causation, if a correlation is clear and consistent, then they typically assume a cause has been found—that there really is some invisible association between the measurements.

Researchers have developed an impressive system for testing these correlations. For the most part, they rely on an abstract measure known as statistical significance, invented by English mathematician Ronald Fisher in the 1920s. This test defines a “significant” result as any data point that would be produced by chance less than 5 percent of the time. While a significant result is no guarantee of truth, it's widely seen as an important indicator of good data, a clue that the correlation is not a coincidence.



Photo: Mauricio Alejo

But here's the bad news: The reliance on correlations has entered an age of diminishing returns. At least two major factors contribute to this trend. First, all of the easy causes have been found, which means that scientists are now forced to search for ever-subtler correlations, mining that mountain of facts for the tiniest of associations. Is that a new cause? Or just a statistical mistake? The line is getting finer; science is getting harder. Second—and this is the biggy—searching for correlations is a terrible way of dealing with the primary subject of much modern research: those complex networks at the center of life. While correlations help us track the relationship between independent measurements, such as the link between smoking and cancer, they are much less effective at making sense of systems in which the variables cannot be isolated. Such situations require that we understand *every* interaction before we can reliably understand any of them. Given the byzantine nature of biology, this can often be a daunting hurdle, requiring that researchers map not only the complete cholesterol pathway but also the ways in which it is plugged into other pathways. (The neglect of these secondary and even tertiary interactions begins to explain the failure of torcetrapib, which had unintended effects on blood pressure. It also helps explain the success of Lipitor, which seems to have a secondary effect of reducing inflammation.) Unfortunately, we often shrug off this dizzying intricacy, searching instead for the simplest of correlations. It's the cognitive equivalent of bringing a knife to a gunfight.

These troubling trends play out most vividly in the drug industry. Although modern pharmaceuticals are supposed to represent the practical payoff of basic research, the R&D to discover a promising new compound now costs about 100 times more (in inflation-adjusted dollars) than it did in 1950. (It also takes nearly three times as long.) This trend shows no sign of letting up: Industry forecasts suggest that once failures are taken into account, the average cost per approved molecule will top \$3.8 billion by 2015. What's worse, even these "successful" compounds don't seem to be worth the investment. According to one internal estimate, approximately 85 percent of new prescription drugs approved by European regulators provide little to no new benefit. We are witnessing Moore's law in reverse.

This returns us to cholesterol, a compound whose scientific history reflects our tortured relationship with causes. At first, cholesterol was entirely bad; the correlations linked high levels of the substance with plaque. Years later, we realized that there were multiple kinds and that only LDL was bad. Then it became clear that HDL was more important than LDL, at least according to correlational studies and animal models. And now we don't really know what matters, since raising HDL levels with torcetrapib doesn't seem to help. Although we've mapped every known part of the chemical pathway, the causes that matter are still nowhere to be found. If this is progress, it's a peculiar kind.

**Back pain** is an epidemic. The numbers are sobering: There's an 80 percent chance that, at some point in your life, you'll suffer from it. At any given time, about 10 percent of Americans are completely incapacitated by their lumbar regions, which is why back pain is the second most frequent reason people seek medical care, after general checkups. And all this treatment is expensive: According to a recent study in *The Journal of the*

*American Medical Association*, Americans spend nearly \$90 billion every year treating back pain, which is roughly equivalent to what we spend on cancer.

When doctors began encountering a surge in patients with lower back pain in the mid-20th century, as I reported for my 2009 book *How We Decide*, they had few explanations. The lower back is an exquisitely complicated area of the body, full of small bones, ligaments, spinal discs, and minor muscles. Then there's the spinal cord itself, a thick cable of nerves that can be easily disturbed. There are so many moving parts in the back that doctors had difficulty figuring out what, exactly, was causing a person's pain. As a result, patients were typically sent home with a prescription for bed rest.

This treatment plan, though simple, was still extremely effective. Even when nothing was done to the lower back, about 90 percent of people with back pain got better within six weeks. The body healed itself, the inflammation subsided, the nerve relaxed.

Over the next few decades, this hands-off approach to back pain remained the standard medical treatment. That all changed, however, with the introduction of magnetic resonance imaging in the late 1970s. These diagnostic machines use powerful magnets to generate stunningly detailed images of the body's interior. Within a few years, the MRI machine became a crucial diagnostic tool.

The view afforded by MRI led to a new causal story: Back pain was the result of abnormalities in the spinal discs, those supple buffers between the vertebrae. The MRIs certainly supplied bleak evidence: Back pain was strongly correlated with seriously degenerated discs, which were in turn thought to cause inflammation of the local nerves. Consequently, doctors began administering epidurals to quiet the pain, and if it persisted they would surgically remove the damaged disc tissue.

But the vivid images were misleading. It turns out that disc abnormalities are typically not the cause of chronic back pain. The presence of such abnormalities is just as likely to be correlated with the absence of back problems, as a 1994 study published in *The New England Journal of Medicine* showed. The researchers imaged the spinal regions of 98 people with no back pain. The results were shocking: Two-thirds of normal patients exhibited "serious problems" like bulging or protruding tissue. In 38 percent of these patients, the MRI revealed multiple damaged discs. Nevertheless, none of these people were in pain. The study concluded that, in most cases, "the discovery of a bulge or protrusion on an MRI scan in a patient with low back pain may frequently be coincidental."

Similar patterns appear in a new study by James Andrews, a sports medicine orthopedist. He scanned the shoulders of 31 professional baseball pitchers. Their MRIs showed that 90 percent of them had abnormal cartilage, a sign of damage that would typically lead to surgery. Yet they were all in perfect health.

This is not the way things are supposed to work. We assume that more information will make it easier to find the cause, that seeing the soft tissue of the back will reveal the

source of the pain, or at least some useful correlations. Unfortunately, that often doesn't happen. Our habits of visual conclusion-jumping take over. All those extra details end up confusing us; the more we know, the less we seem to understand.

The only solution for this mental flaw is to deliberately ignore a wealth of facts, even when the facts seem relevant. This is what's happening with the treatment of back pain: Doctors are now encouraged to *not* order MRIs when making diagnoses. The latest clinical guidelines issued by the American College of Physicians and the American Pain Society strongly recommended that doctors “not routinely obtain imaging or other diagnostic tests in patients with nonspecific low back pain.”

And it's not just MRIs that appear to be counterproductive. Earlier this year, John Ioannidis, a professor of medicine at Stanford, conducted an in-depth review of biomarkers in the scientific literature. Biomarkers are molecules whose presence, once detected, are used to infer illness and measure the effect of treatment. They have become a defining feature of modern medicine. (If you've ever had your blood drawn for lab tests, you've undergone a biomarker check. Cholesterol is a classic biomarker.) Needless to say, these tests depend entirely on our ability to perceive causation via correlation, to link the fluctuations of a substance to the health of the patient.

In his resulting paper, published in *JAMA*, Ioannidis looked at only the most highly cited biomarkers, restricting his search to those with more than 400 citations in the highest impact journals. He identified biomarkers associated with cardiovascular problems, infectious diseases, and the genetic risk of cancer. Although these causal stories had initially triggered a flurry of interest—several of the biomarkers had already been turned into popular medical tests—Ioannidis found that the claims often fell apart over time. In fact, 83 percent of supposed correlations became significantly weaker in subsequent studies.

Consider the story of homocysteine, an amino acid that for several decades appeared to be linked to heart disease. The original paper detecting this association has been cited 1,800 times and has led doctors to prescribe various B vitamins to reduce homocysteine. However, a study published in 2010—involving 12,064 volunteers over seven years—showed that the treatment had no effect on the risk of heart attack or stroke, despite the fact that homocysteine levels were lowered by nearly 30 percent.

The larger point is that we've constructed our \$2.5 trillion health care system around the belief that we can find the underlying causes of illness, the invisible triggers of pain and disease. That's why we herald the arrival of new biomarkers and get so excited by the latest imaging technologies. If only we knew more and could see further, the causes of our problems would reveal themselves. But what if they don't?

The failure of this drug in particular has not ended the development of new cholesterol medications. The potential market for them is simply too huge.



**The failure of torcetrapib** has not ended the development of new cholesterol medications—the potential market is simply too huge. Although the compound is a sobering reminder that our causal beliefs are defined by their oversimplifications, that even the best-understood systems are still full of surprises, scientists continue to search for the magic pill that will make cardiovascular disease disappear. Ironically, the latest hyped treatment, a drug developed by Merck called anacetrapib, inhibits the exact same protein as torcetrapib. The initial results of the clinical trial, which were made public in November 2010, look promising. Unlike its chemical cousin, this compound doesn't appear to raise systolic blood pressure or cause heart attacks. (A much larger clinical trial is under way to see whether the drug saves lives.) Nobody can conclusively explain why these two closely related compounds trigger such different outcomes or why, according to a 2010 analysis, high HDL levels might actually be dangerous for some people. We know so much about the cholesterol pathway, but we never seem to know what matters.

Chronic back pain also remains a mystery. While doctors have long assumed that there's a valid correlation between pain and physical artifacts—a herniated disc, a sheared muscle, a pinched nerve—there's a growing body of evidence suggesting the role of seemingly unrelated factors. For instance, a recent study published in the journal *Spine* concluded that minor physical trauma had virtually no relationship with disabling pain. Instead, the researchers found that a small subset of “nonspinal factors,” such as depression and smoking, were most closely associated with episodes of serious pain. We keep trying to fix the back, but perhaps the back isn't what needs fixing. Perhaps we're searching for causes in the wrong place.

The same confusion afflicts so many of our most advanced causal stories. Hormone replacement therapy was supposed to reduce the risk of heart attack in postmenopausal women—estrogen prevents inflammation in blood vessels—but a series of recent clinical trials found that it did the opposite, at least among older women. (Estrogen therapy was also supposed to ward off Alzheimer's, but that doesn't seem to work, either.) We were told that vitamin D supplements prevented bone loss in people with multiple sclerosis and that vitamin E supplements reduced cardiovascular disease—neither turns out to be true.

It would be easy to dismiss these studies as the inevitable push and pull of scientific progress; some papers are bound to get contradicted. What's remarkable, however, is just how common such papers are. One study, for instance, analyzed 432 different claims of genetic links for various health risks that vary between men and women. Only one of these claims proved to be consistently replicable. Another meta review, meanwhile, looked at the 49 most-cited clinical research studies published between 1990 and 2003. Most of these were the culmination of years of careful work. Nevertheless, more than 40 percent of them were later shown to be either totally wrong or significantly incorrect. The details always change, but the story remains the same: We think we understand how something works, how all those shards of fact fit together. But we don't.

Given the increasing difficulty of identifying and treating the causes of illness, it's not surprising that some companies have responded by abandoning entire fields of research. Most recently, two leading drug firms, AstraZeneca and GlaxoSmithKline, announced

that they were scaling back research into the brain. The organ is simply too complicated, too full of networks we don't comprehend.

David Hume referred to causality as “the cement of the universe.” He was being ironic, since he knew that this so-called cement was a hallucination, a tale we tell ourselves to make sense of events and observations. No matter how precisely we knew a given system, Hume realized, its underlying causes would always remain mysterious, shadowed by error bars and uncertainty. Although the scientific process tries to make sense of problems by isolating every variable—imagining a blood vessel, say, if HDL alone were raised—reality doesn't work like that. Instead, we live in a world in which everything is knotted together, an impregnable tangle of causes and effects. Even when a system is dissected into its basic parts, those parts are still influenced by a whirligig of forces we can't understand or haven't considered or don't think matter. Hamlet was right: There really are more things in heaven and Earth than are dreamt of in our philosophy.

This doesn't mean that nothing can be known or that every causal story is equally problematic. Some explanations clearly work better than others, which is why, thanks largely to improvements in public health, the average lifespan in the developed world continues to increase. (According to the Centers for Disease Control and Prevention, things like clean water and improved sanitation—and not necessarily advances in medical technology—accounted for at least 25 of the more than 30 years added to the lifespan of Americans during the 20th century.) Although our reliance on statistical correlations has strict constraints—which limit modern research—those correlations have still managed to identify many essential risk factors, such as smoking and bad diets.

And yet, we must never forget that our causal beliefs are defined by their limitations. For too long, we've pretended that the old problem of causality can be cured by our shiny new knowledge. If only we devote more resources to research or dissect the system at a more fundamental level or search for ever more subtle correlations, we can discover how it all works. But a cause is not a fact, and it never will be; the things we can see will always be bracketed by what we cannot. And this is why, even when we know everything about everything, we'll still be telling stories about why it happened. It's mystery all the way down.

Learn why all science experiments have error, how to calculate it, and the sources and types of errors you should report. The three main categories of errors are systematic errors, random errors, and personal errors. Here's what these types of errors are and common examples. Systematic Errors. Systematic error affects all the measurements you take. All of these errors will be in the same direction (greater than or less than the true value) and you can't compensate for them by taking additional data. Examples of Systematic Errors. If you forget to calibrate a balance or you're off a bit in the calibration, all mass measurements will be high/low by the same amount. Yet, according to a recent Gartner study, 85% of data science projects fail. Why? More to the point, how can your company be in the 15% (Walker, 2017)? This successful multi-year project nearly exploded on the launch pad for one of the same reasons data science projects fail " they erroneously assumed that a programmer skilled in a few languages and tools could: (a) do everything with those couple tools; and, (b) figure out all the other elements " successfully identifying requirements, technical strategies, statistics, user experience, quality assurance, training, and delivery. Main topics: the failure of drugs in clinical trials, diminishing returns to pharmaceutical research, doctors over-treating, and Humean causality-correlation distinction, with some Ioannidis mixed through-out. See also "Why epidemiology will not correct itself". In completely unrelated news, Nick Bostrom is stepping down from IEET's Chairman of the Board.