

Cancer Research Is Broken: There's a replication crisis in biomedicine—and no one even knows how deep it runs.

By Daniel Engber

When a cancer study ends up at the wrong conclusion, people die and suffer, and a multibillion-dollar industry of treatment loses money, too. The U.S. government spends \$5 billion every year on cancer research; charities and private firms add billions more. Yet the return on this investment—in terms of lives saved, suffering reduced—has long been disappointing: Cancer death rates are drifting downward in the past 20 years, but not as quickly as we'd hoped. Even as the science makes incremental progress, it feels as though we're going in a circle.

That's a "hackable problem," says Silicon Valley pooh-bah Sean Parker, who last week announced his founding of a \$250 million institute for research into cancer immunotherapy, an old idea that's come around again in recent years. "As somebody who has spent his life as an entrepreneur trying to pursue kind of rapid, disruptive changes," Parker said, "I'm impatient."

Many science funders share Parker's antsiness over all the waste of time and money. In February, the White House announced its plan to put \$1 billion toward a similar objective—a "Cancer Moonshot" aimed at making research more techy and efficient. But recent studies of the research enterprise reveal a more confounding issue, and one that won't be solved with bigger grants and increasingly disruptive attitudes. The deeper problem is that much of cancer research in the lab—maybe even most of it—simply can't be trusted. The data are corrupt. The findings are unstable. The science doesn't work.

The resulting waste amounts to more than \$28 billion: That's two dozen Cancer Moonshots misfired in every single year.

In other words, we face a replication crisis in the field of biomedicine, not unlike the one we've seen in psychology but with far more dire implications. Sloppy data analysis, contaminated lab materials, and poor experimental design all contribute to the problem. Last summer, Leonard P. Freedman, a scientist who worked for years in both academia and big pharma, published a paper with two colleagues on "the economics of reproducibility in preclinical research." After reviewing the estimated prevalence of each of these flaws and fault-lines in biomedical literature, Freedman and his co-authors guessed that fully half of all results rest on shaky ground, and might not be replicable in other labs. These cancer studies don't merely fail to find a cure; they might not offer any useful data whatsoever. Given current U.S. spending habits, the resulting waste amounts to more than \$28 billion. That's two dozen Cancer Moonshots misfired in every single year. That's 100 squandered internet tycoons.

How could this be happening? At first glance it would seem medical research has a natural immunity to the disease of irreproducible results. Other fields, such as psychology, hold a more tenuous connection to our lives. When a social-science theory turns to be misguided, we have only to update our understanding of the human mind—a shift of attitude, perhaps, as opposed to one of practice. The real-world stakes are low enough that strands of falsehood might sustain themselves throughout the published literature without having too much impact. But when a cancer study ends up at the wrong conclusion—and an errant strand is snipped—people die and suffer, and a multibillion-dollar industry of treatment loses money, too. I always figured that this feedback would provide a self-corrective loop, a way for the invisible hands of human health and profit motive to guide the field away from bad technique.

Alas, the feedback loop doesn't seem to work so well, and without some signal to correct them, biologists get stuck in their bad habits, favoring efficiency in publication over the value of results. They also face a problem more specific to their research: The act of reproducing biomedical experiments—I mean, just attempting to obtain the same result—takes enormous time and money, far more than would be required for, say, studies in psychology. That makes it very hard to diagnose the problem of reproducibility in cancer research and understand its scope and symptoms. If we can't easily test the literature for errors, then how are we supposed to fix it up?

When cancer research does get tested, it's almost always by a private research lab. Pharmaceutical and biotech businesses have the money and incentive to proceed—but these companies mostly keep their findings to themselves. (That's another break in the feedback loop of self-correction.) In 2012, the former head of cancer research at Amgen, Glenn Begley, brought wide attention to this issue when he decided to go public with his findings in a piece for *Nature*. Over a 10-year stretch, he said, Amgen's scientists had tried to replicate the findings of 53 "landmark" studies in cancer biology. Just six of them came up with positive results.

Begley blames these failures on some systematic problems in the literature, not just in cancer research but all of biomedicine. He says that preclinical work—the basic science often done by government-funded, academic scientists—tends to be quite slipshod. Investigators fail to use controls; or they don't blind themselves to study groups; or they selectively report their data; or they skip important steps, such as testing their reagents. Begley's broadside came as no surprise to those in the industry. In 2011, a team from Bayer had reported that only 20 to 25 percent of the studies they tried to reproduce came to results "completely in line" with those of the original publications. There's even a rule of thumb among venture capitalists, the authors noted, that at least half of published studies, even those from the very best journals, will not work out the same when conducted in an industrial lab.

An international effort to pool the findings from this hidden, private research on reliability could help us to assess the broader problem. The Reproducibility Project for Cancer Biology, started in 2013 with money from the Laura and John Arnold Foundation, should be even better. The team behind the project chose 50 highly influential papers published between 2010 and 2012, and then set out to work in concert with the authors of each one, so as to reconstruct the papers' most important, individual experiments. Once everyone agreed upon the details—and published them in a peer-reviewed journal—the team farmed out experiments to unbiased, commercial research groups. (The contracts are being handled through the Science Exchange, a Silicon Valley startup that helps to allocate research tasks to a network of more than 900 private labs.)

Problems and delays occurred at every level. The group started with about \$2 million, says Brian Nosek, a psychologist and advocate for replication research, who helped to launch the Reproducibility Project for Cancer Biology as well as an earlier, analogous one for psychology. The psychology project, which began in late 2011 and was published last summer, reproduced studies from 100 different papers on the basis of a \$250,000 grant and lots of volunteered time from the participants. The cancer biology project, on the other hand, has only tried to cover half that many studies, on a budget eight times larger—and even that proved to be too much. Last summer the group was forced to scale back its plan from evaluating 50 papers to looking at just 35 or 37.

It took months and months just to negotiate the paperwork, says Elizabeth Iorns, a member of the project team and founder of Science Exchange. Many experiments required contracts, called "material transfer agreements," that would allow one institution to share its cells, animals, or bits of DNA with another research group. Iorns recalls that it took a full year of back-and-forth communication to work out just one of these agreements.

For some experiments, the original materials could not be shared, red tape notwithstanding, because they were simply gone or corrupted in some way. That meant the replicating labs would have to recreate the materials themselves—an arduous undertaking. Iorns said one experiment called for the creation of a quadruple-transgenic mouse, i.e. one with its genome modified in four specific ways. “It would take literally years and years to produce them,” she said. “We decided that it was not going to happen.”

Then there’s the fact that the heads of many labs have little sense of how, exactly, their own experiments were carried out. In many cases, a graduate student or post-doc did most of the work, and then moved on to another institution. To reconstruct the research, then, someone had to excavate and analyze the former student or post-doc’s notes—a frustrating, time-consuming task. “A lot of time we don’t know what reagents the original lab used,” said Tim Errington, the project’s manager, “and the original lab doesn’t know, either.”

I talked to one researcher, Cory Johannessen of the Broad Institute in Cambridge, Massachusetts, whose 2010 paper on the development of drug-resistance in cancer cells has been selected for replication. Most of the actual work was done in 2008, he told me. While he said that he was glad to have been chosen for the project, the process turned into a nightmare. His present lab is just down the hall from the one in which he’d done the research, and even so he said “it was a huge, huge amount of effort” to sort through his old materials. “If I were on the other side of the country, I would have said, ‘I’m sorry, I can’t help.’” Reconstructing what he’d done eight years ago meant digging up old notebooks and finding details as precise as how many cells he’d seeded in each well of a culture plate. “It felt like coming up with a minute-by-minute protocol,” he said.

The Reproducibility Project also wants to use the same materials and reagents as the original researchers, even buying from the same suppliers, when possible. A few weeks ago, the team published its official plan to reproduce six experiments from Johannessen’s paper; the first one alone calls for about three dozen materials and tools, ranging from the trivial—“6-well plates, original brand not specified”—to key cell lines.

Johannessen’s experience, and the project’s as a whole, illustrate that the crisis of “reproducibility” has a double meaning. In one sense, it’s a problem of results: Can a given finding be repeated in another lab? Does the finding tell us something true about the world? In another sense, though, it’s a problem of methodology: Can a given experiment even be repeated in another lab, whatever its results might be? If there’s no way to reproduce experiments, there’s no way to know if we can reproduce results.

Research on the research literature shows the wideness of the methodology problem. Even basic facts about experiments—essential steps that must be followed while recreating research—are routinely omitted from published papers. A 2009 survey of animal studies found that just 60 percent included information about the number, strain, sex, age, or weight of the animals used. Another survey, published in 2013, looked at several hundred journal articles which together made reference to more than 1,700 different laboratory materials. According to the authors, only about half of these materials could be identified by reading the original papers.

One group of researchers tried to reach out to the investigators behind more than 500 original research papers published between 1991 and 2011, and found that just one-quarter of those authors said they had their data. (Though not all were willing to share.) Another 25 percent of the authors were simply unreachable—the research team could not find a working email address for them.

Not all these problems are unique to biomedicine. Brian Nosek points out that in most fields, career advancement comes with publishing the most papers and the flashiest papers, not the most well-documented ones. That means that when it comes to getting ahead, it’s not really in the interest of any researchers—biologists and psychologists alike—to be comprehensive in the reporting of their data and procedures. And for every point that one could make about the specific problems with reproducing biology experiments—the trickiness of identifying biological reagents, or working out complicated protocols—Nosek offers an analogy from his field. Even a behavioral study of local undergraduate volunteers may require subtle calibrations, careful delivery of instructions, and attention to seemingly trivial factors such as the time of day. I thought back to what the social psychologist Roy Baumeister told me about his own work, that there is “a craft to running experiments,” and that this craft is sometimes bungled in attempted replications.

That may be so, but I’m still convinced that psychology has a huge advantage over cancer research, when it comes to self-diagnosis. You can see it in the way each field has responded to its replication crisis. Some psychology labs are now working to validate and replicate their own research before it’s published. Some psychology journals are requiring researchers to announce their research plans and hypotheses ahead of time, to help prevent bias. And though its findings have been criticized, the Reproducibility Project for Psychology has already been completed. (This openness to dealing with the problem may explain why the crisis in psychology has gotten somewhat more attention in the press.)

The biologists, in comparison, have been reluctant or unable to pursue even very simple measures of reform. Leonard Freedman, the lead author of the paper on the economics of irreproducibility, has been pushing very hard for scientists to pay attention to the cell lines that they use in research. These common laboratory tools are often contaminated with hard-to-see bacteria, or else with other, unrelated lines of cells. One survey found such problems may affect as many as 36 percent of the cell lines used in published papers. Freedman notes that while there is a simple way to test a cell line for contamination—a genetic test that costs about a hundred bucks—it’s almost never used. Some journals recommend the test, but almost none require it. “Deep down, I think they’re afraid to make the bar too high,” he said.

Likewise, Elizabeth Iorns’ Science Exchange launched a “Reproducibility Initiative” in 2012, so that biomedical researchers could use her network to validate their own findings in an independent laboratory. Four years later, she says that not a single lab has taken that opportunity. “We didn’t push it very hard,” she said, explaining that any researchers who paid to reproduce their research might be accused of having “misused” their grant funding. (That problem will persist until the people giving out the grants—including those in government—agree to make such validations a priority.) Iorns now calls the whole thing an “awareness-building experiment.”

Projects like the ones led by Brian Nosek and Elizabeth Iorns may push cancer researchers to engage more fully with the problems in their field. Whatever number the Reproducibility Project ends up putting on the cancer research literature—I mean, whatever percentage of the published findings the project manages to reproduce—may turn out to be less important than the finding it has made already: Namely, that the process is itself a hopeless slog. We’ll never understand the problems with cancer studies, let alone figure out a way to make them “hackable,” until we’ve figured out a way to make them reproducible.

(Accessed on 9.19.17 at http://www.slate.com/articles/health_and_science/future_tense/2016/04/biomedicine_facing_a_worse_replication_crisis_than_the_one_plaguing_psychology.html)