

Dow Jones Reprints: This copy is for your personal, non-commercial use only. To order presentation-ready copies for distribution to your colleagues, clients or customers, use the Order Reprints tool at the bottom of any article or visit www.djreprints.com

See a sample reprint in PDF format.

Order a reprint of this article now

THE WALL STREET JOURNAL.
WSJ.com

HEALTH INDUSTRY | DECEMBER 2, 2011

Scientists' Elusive Goal: Reproducing Study Results

By GAUTAM NAIK

Two years ago, a group of Boston researchers published a study describing how they had destroyed cancer tumors by targeting a protein called STK33. Scientists at biotechnology firm [Amgen](#) Inc. quickly pounced on the idea and assigned two dozen researchers to try to repeat the experiment with a goal of turning the findings into a drug.



WSJ's Gautam Naik has details of challenges scientists face in reproducing claims made by medical journals. Photo: Sandy Huffaker/The New York Times

It proved to be a waste of time and money. After six months of intensive lab work, Amgen found it couldn't replicate the results and scrapped the project.

"I was disappointed but not surprised," says Glenn Begley, vice president of research at Amgen of Thousand Oaks, Calif. "More often than not, we are unable to reproduce findings" published by researchers in journals.

This is one of medicine's dirty secrets: Most results, including those that appear in top-flight peer-reviewed journals, can't be reproduced.

"It's a very serious and disturbing issue because it obviously misleads people" who implicitly trust findings published in a respected peer-reviewed journal, says Bruce Alberts, editor of *Science*. On Friday, the U.S. journal is devoting a large chunk of its Dec. 2 issue to the problem of scientific replication.

Reproducibility is the foundation of all modern research, the standard by which scientific claims are evaluated. In the U.S. alone, biomedical research is a \$100-billion-year enterprise. So when published medical findings can't be validated by others, there are major consequences.

Drug manufacturers rely heavily on early-stage academic research and can waste millions of dollars on products if the

original results are later shown to be unreliable. Patients may enroll in clinical trials based on conflicting data, and sometimes see no benefits or suffer harmful side effects.

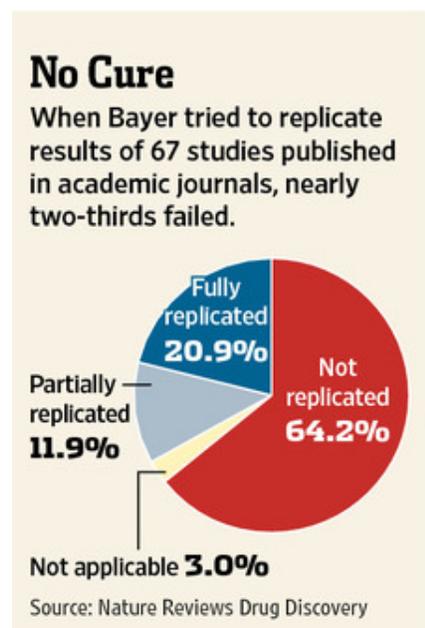
There is also a more insidious and pervasive problem: a preference for positive results.

Unlike pharmaceutical companies, academic researchers rarely conduct experiments in a "blinded" manner. This makes it easier to



Bayer

Researchers at Bayer's labs often find their experiments fail to match claims made in the scientific literature.



cherry-pick statistical findings that support a positive result. In the quest for jobs and funding, especially in an era of economic malaise, the growing army of scientists need more successful experiments to their name, not failed ones. An explosion of scientific and academic journals has added to the pressure.

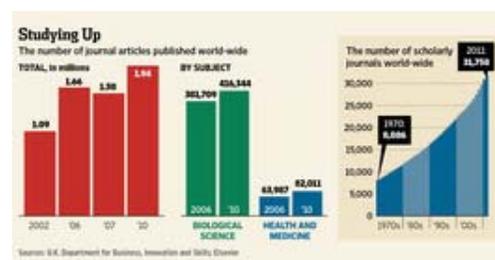
When it comes to results that can't be replicated, Dr. Alberts says the increasing intricacy of experiments may be largely to blame. "It has to do with the complexity of biology and the fact that methods [used in labs] are getting more sophisticated," he says.

It is hard to assess whether the reproducibility problem has been getting worse over the years; there are some signs suggesting it could be. For example, the success rate of Phase 2 human trials—where a drug's efficacy is measured—fell to 18% in 2008-2010 from 28% in 2006-2007, according to a global analysis published in the journal *Nature Reviews* in May.

"Lack of reproducibility is one element in the decline in Phase 2 success," says Khusru Asadullah, a Bayer AG research executive.

In September, Bayer published a study describing how it had halted nearly two-thirds of its early drug target projects because in-house experiments failed to match claims made in the literature.

The German pharmaceutical company says that none of the claims it attempted to validate were in papers that had been retracted or were suspected of being flawed. Yet, even the data in the most prestigious journals couldn't be confirmed, Bayer said.



In 2008, [Pfizer Inc.](#) made a high-profile bet, potentially worth more than \$725 million, that it could turn a 25-year-old Russian cold medicine into an effective drug for Alzheimer's disease.

The idea was promising. Published by the journal *Lancet*, data from researchers at Baylor College of Medicine and elsewhere suggested that the drug, an antihistamine called Dimebon, could improve symptoms in Alzheimer's patients. Later findings, presented by researchers at the University of California

Los Angeles at a Chicago conference, showed that the drug appeared to prevent symptoms from worsening for up to 18 months.

"Statistically, the studies were very robust," says David Hung, chief executive officer of Medivation Inc., a San Francisco biotech firm that sponsored both studies.

In 2010, Medivation along with Pfizer released data from their own clinical trial for Dimebon, involving nearly 600 patients with mild to moderate Alzheimer's disease symptoms. The companies said they were unable to reproduce the *Lancet* results. They also indicated they had found no statistically significant difference between patients on the drug versus the inactive placebo.

Pfizer and Medivation have just completed a one-year study of Dimebon in over 1,000 patients, another effort to see if the drug could be a potential treatment for Alzheimer's. They expect to announce the results in coming months.

Scientists offer a few theories as to why duplicative results may be so elusive. Two different labs can use slightly different equipment or materials, leading to divergent results. The more variables there are in an experiment, the more likely it is that small, unintended errors will pile up and swing a lab's conclusions one way or the other.

And, of course, data that have been rigged, invented or fraudulently altered won't stand up to future scrutiny.

According to a report published by the U.K.'s Royal Society, there were 7.1 million researchers working globally across all scientific fields—academic and corporate—in 2007, a 25% increase from five years earlier.

From the Archives

Mistakes in Scientific Studies Surge
8/10/2011

"Among the more obvious yet unquantifiable reasons, there is immense competition among laboratories and a pressure to publish," wrote Dr. Asadullah and others from Bayer, in their September paper. "There is also a bias toward publishing positive results, as it is easier to get positive results accepted in

good journals."

Science publications are under pressure, too. The number of research journals has jumped 23% between 2001 and 2010, according to Elsevier, which has analyzed the data. Their proliferation has ratcheted up competitive pressure on even elite journals, which can generate buzz by publishing splashy papers, typically containing positive findings, to meet the demands of a 24-hour news cycle.

Dr. Alberts of Science acknowledges that journals increasingly have to strike a balance between publishing studies "with broad appeal," while making sure they aren't hyped.

Drugmakers also have a penchant for positive results. A 2008 study published in the journal PLoS Medicine by researchers at the University of California San Francisco looked at data from 33 new drug applications submitted between 2001 and 2002 to the U.S. Food and Drug Administration. The agency requires drug companies to provide all data from clinical trials. However, the authors found that a quarter of the trial data—most of it unfavorable—never got published because the companies never submitted it to journals.

The upshot: doctors who end up prescribing the FDA-approved drugs often don't get to see the unfavorable data.

"I would say that selectively publishing data is unethical because there are human subjects involved," says Lisa Bero of UCSF and co-author of the PLoS Medicine study.

In an email statement, a spokeswoman for the FDA said the agency considers all data it is given when reviewing a drug but "does not have the authority to control what a company chooses to publish."

Venture capital firms say they, too, are increasingly encountering cases of nonrepeatable studies, and cite it as a key reason why they are less willing to finance early-stage projects. Before investing in very early-stage research, Atlas Ventures, a venture-capital firm that backs biotech companies, now asks an outside lab to validate any experimental data. In about half the cases the findings can't be reproduced, says Bruce Booth, a partner in Atlas' Life Sciences group.

There have been several prominent cases of nonreproducibility in recent months. For example, in September, the journal Science partially retracted a 2009 paper linking a virus to chronic fatigue syndrome because several labs couldn't replicate the published results. The partial retraction came after two of the 13 study authors went back to the blood samples they analyzed from chronic-fatigue patients and found they were contaminated.

Some studies can't be redone for a more prosaic reason: the authors won't make all their raw data available to rival scientists.

John Ioannidis of Stanford University recently attempted to reproduce the findings of 18 papers published in the respected journal Nature Genetics. He noted that 16 of these papers stated that the underlying "gene expression" data for the studies were publicly available.

But the supplied data apparently weren't detailed enough, and results from 16 of the 18 major papers couldn't fully be reproduced by Dr. Ioannidis and his colleagues. "We have to take it [on faith] that the findings are OK," said Dr. Ioannidis, an epidemiologist who studies the credibility of medical research.

Veronique Kiermer, an editor at Nature, says she agrees with Dr. Ioannidis' conclusions, noting that the findings have prompted the journal to be more cautious when publishing large-scale genome analyses.

When companies trying to find new drugs come up against the nonreproducibility problem, the repercussions can be significant.

A few years ago, several groups of scientists began to seek out new cancer drugs by targeting a protein called KRAS. The KRAS protein transmits signals received on the outside of a cell to its interior and is therefore crucial for regulating cell growth. But when certain mutations occur, the signaling can become continuous. That triggers excess growth such as tumors.

The mutated form of KRAS is believed to be responsible for more than 60% of pancreatic cancers and half of colorectal cancers. It has also been implicated in the growth of tumors in many other organs, such as the lung.

So scientists have been especially keen to impede KRAS and, thus, stop the constant signaling that leads to tumor growth.

In 2008, researchers at Harvard Medical School used cell-culture experiments to show that by inhibiting another protein, STK33, they could prevent the growth of tumor cell lines driven by the malfunctioning KRAS.

The finding galvanized researchers at Amgen, who first heard about the experiments at a scientific conference. "Everyone was trying to do this," recalls Dr. Begley of Amgen, which derives nearly half of its revenues from cancer drugs and related treatments. "It was a really big deal."

When the Harvard researchers published their results in the prestigious journal *Cell*, in May 2009, Amgen moved swiftly to capitalize on the findings.

At a meeting in the company's offices in Thousand Oaks, Calif., Dr. Begley assigned a group of Amgen researchers the task of identifying small molecules that might inhibit STK33. Another team got a more basic job: reproduce the Harvard data.

"We're talking about hundreds of millions of dollars in downstream investments" if the approach works," says Dr. Begley. "So we need to be sure we're standing on something firm and solid."

But over the next few months, Dr. Begley and his team got increasingly disheartened. Amgen scientists, it turned out, couldn't reproduce any of the key findings published in *Cell*.

For example, there was no difference in the growth of cells where STK33 was largely blocked, compared with a control group of cells where STK33 wasn't blocked.

What could account for the irreproducibility of the results?

"In our opinion there were methodological issues" in Amgen's approach that could have led to the different findings, says Claudia Scholl, one of the lead authors of the original *Cell* paper.

Dr. Scholl points out, for example, that Amgen used a different reagent to suppress STK33 than the one reported in *Cell*. Yet, she acknowledges that even when slightly different reagents are used, "you should be able to reproduce the results."

Now a cancer researcher at the University Hospital of Ulm in Germany, Dr. Scholl says her team has reproduced the original *Cell* results multiple times, and continues to have faith in STK33 as a cancer target.

Amgen, however, killed its STK33 program. In September, two dozen of the firm's scientists published a paper in the journal *Cancer Research* describing their failure to reproduce the main *Cell* findings.

Dr. Begley suggests that academic scientists, like drug companies, should perform more experiments in a "blinded" manner to reduce any bias toward positive findings. Otherwise, he says, "there is a human desire to get

the results your boss wants you to get."

Adds Atlas' Mr. Booth: "Nobody gets a promotion from publishing a negative study."

Write to Gautam Naik at gautam.naik@wsj.com

Copyright 2011 Dow Jones & Company, Inc. All Rights Reserved

This copy is for your personal, non-commercial use only. Distribution and use of this material are governed by our [Subscriber Agreement](#) and by copyright law. For non-personal use or to order multiple copies, please contact Dow Jones Reprints at 1-800-843-0008 or visit www.djreprints.com