

NOT AS EASY AS 1, 2, 3:

Groundbreaking research gets published in top journal // Replication study finds less promising result // Third study can't reproduce results // If such studies are **done at all**

# The Problem of Replication

■ BY LINDA KESLAR // PHOTOGRAPHS BY SAM KAPLAN

**I**n 2012, when the journal *Science* published a study by a group of Cleveland researchers touting a cancer drug's success against Alzheimer's disease in mice, it looked like a potential breakthrough in a field long marked by setbacks and failures. As many as 5.4 million Americans have Alzheimer's, and by one estimate, there's a new case diagnosed every 69 seconds. While existing treatments can alleviate symptoms, there's nothing to stop the progression of the ultimately fatal disease.

But in this study, researchers from Case Western Reserve University School of Medicine described how a drug called bexarotene appeared to clear the buildup of amyloid-beta plaques—protein deposits in the brain that are hallmarks of Alzheimer's—in mice genetically engineered to have a condition similar to the human disease. The drug also seemed to improve the mice's memory, cognitive abilities and social behavior. “The plaque reduction was an astounding finding,” says Sangram Sisodia, director of the Center for Molecular Neurobiology at the University of Chicago.

No other drug had attacked the plaques so rapidly. Bexarotene, marketed under the brand name Targretin, belongs to the retinoid class of drugs, related to vitamin A, that had been approved by the Food and Drug Administration in 1999 to treat a type of skin cancer called cutaneous T-cell lymphoma. According to the study in *Science*, bexarotene activated a gene that boosts production of apolipoprotein E, or ApoE, which can help break down amyloid-beta. Alzheimer's patients can't

produce enough ApoE to prevent the protein from creating deposits. The drug also seemed to rouse a cellular mechanism that consumes the plaques.

Sisodia was one of many scientists who rushed to replicate the original study. “I think we all went back to our labs and tried to confirm these promising findings by repeating the initial experiments,” he says. Rudolph Tanzi, director of the genetics and aging research unit at Massachusetts General Hospital, also sought to repeat the research, spurred in part by physicians whose patients had heard about the apparent breakthrough and wanted to start taking bexarotene immediately. That was such a common response that in August 2012, *The New England Journal of Medicine* published an article warning physicians to wait for evidence from human clinical trials. “The results in mice certainly didn't guarantee the drug would work in people,” says Tanzi. “And although it's safe for cancer treatment, it can have severe side effects.”

But before testing whether the drug could help human Alzheimer's patients, scientists needed to make sure it worked in mice—and while some labs were able to reproduce the drug's effect on memory and cognition, the plaque reduction effect couldn't be replicated. In four technical comments published in *Science* last May, several teams of independent researchers, including those co-authored by Sisodia and Tanzi, reported that their replication attempts showed no effect on plaque levels in lab mice treated with bexarotene. Months later, in August, a paper in the journal *Molecular Neurodegeneration* reported that researchers from Johns Hopkins had also failed to

replicate either the plaque reduction or memory and cognition effects found in the Case Western research.

For anyone hoping to see progress in the fight against Alzheimer's, the failure of the follow-up research was disappointing. But it also points to the necessity and challenges of independently validating published research findings. While reproducibility is considered a bedrock of scientific discovery, there has been growing concern about the quality of recent studies. "Data reproducibility means that the seminal findings of a paper can be reproduced in any qualified lab that has appropriate resources and expertise," says Lee Ellis, a surgeon and researcher at the University of Texas MD Anderson Cancer Center in Houston. "If you try to reproduce all of the findings in a paper, you're likely to find some divergent outcomes, but the point of the paper should remain the same."

But Ellis and others who have explored these issues have found that medical research, including seemingly groundbreaking work, is reproducible less than half the time. "The unspoken rule is that at least 50% and more like 70% of the studies published even in top-tier academic journals can't be repeated," says Bruce Booth, a partner at Atlas Venture, a venture capital firm in Boston. "Everyone recognizes reproducibility as a big problem," says Elizabeth Iorns, a cancer researcher in Palo Alto, Calif., and chief executive of Science Exchange, an online marketplace for scientific resources and expertise.

Many factors contribute to the low odds of reproducibility. The original experiments may have been poorly designed, or there could be problems with how results were analyzed. The trend may also be a symptom of a scientific community in which the job market and funding are tighter than ever, and researchers must publish or perish, leading to a lack of rigor



in their research. "It is a dysfunctional scientific climate," says Ferric Fang, a professor at the University of Washington School of Medicine and editor-in-chief of the journal *Infection and Immunity*. And because journals favor original research, scientists have little incentive to pursue replicative work.

As intractable as those issues may seem, there are compelling reasons to address them. Pharmaceutical and biotechnology companies depend on academic research when developing new drugs, and erroneous studies waste time and money. "Not being able to rely on research results has made early-stage investing harder," says Booth, who is an advisor to the Reproducibility Initiative, a network launched by Iorns and other scientists to help researchers independently validate study findings. The National Institutes of Health, meanwhile, has established pilot programs to address replication problems, and some leading science journals are raising the bar on their standards

for publication. “Everyone is asking whether this is something we can fix, but it’s clear there are no simple answers,” Fang says.

**R**epeated experimentation has always been a foundation of scientific discovery. In the 17th century, Robert Boyle, considered the first modern chemist, argued that if findings were to be credible and reliable, they had to be based on methods that independent researchers could learn, assess and replicate. Three centuries later, Austrian philosopher Karl Popper, writing in *The Logic of Scientific Discovery* in 1934, asserted that “non-reproducible single occurrences are of no significance to science.”

Yet while few may question the importance of replication, the technology and complexity of scientific experimentation today can make it enormously challenging. “A lot of techniques in my laboratory take a long time to master, and there’s a steep learning curve before we can reproduce even our own results,” says Fang, a microbiologist. “So another lab saying ‘We’re going to repeat the high-energy UV laser footprinting you just did on those nucleoprotein complexes’ is going to find it very daunting—and that’s just one component of the experiment.”

But the growing complexity of research methodologies is hardly the only reason replication is no longer a routine part of scientific discovery. “A big factor is that scientists have strong incentives to introduce new ideas, but weak ones to confirm the validity of old ideas,” says Brian Nosek, a psychologist at the University of Virginia. “Innovative findings produce rewards of publication, employment and tenure. Replicated findings produce a shrug.”

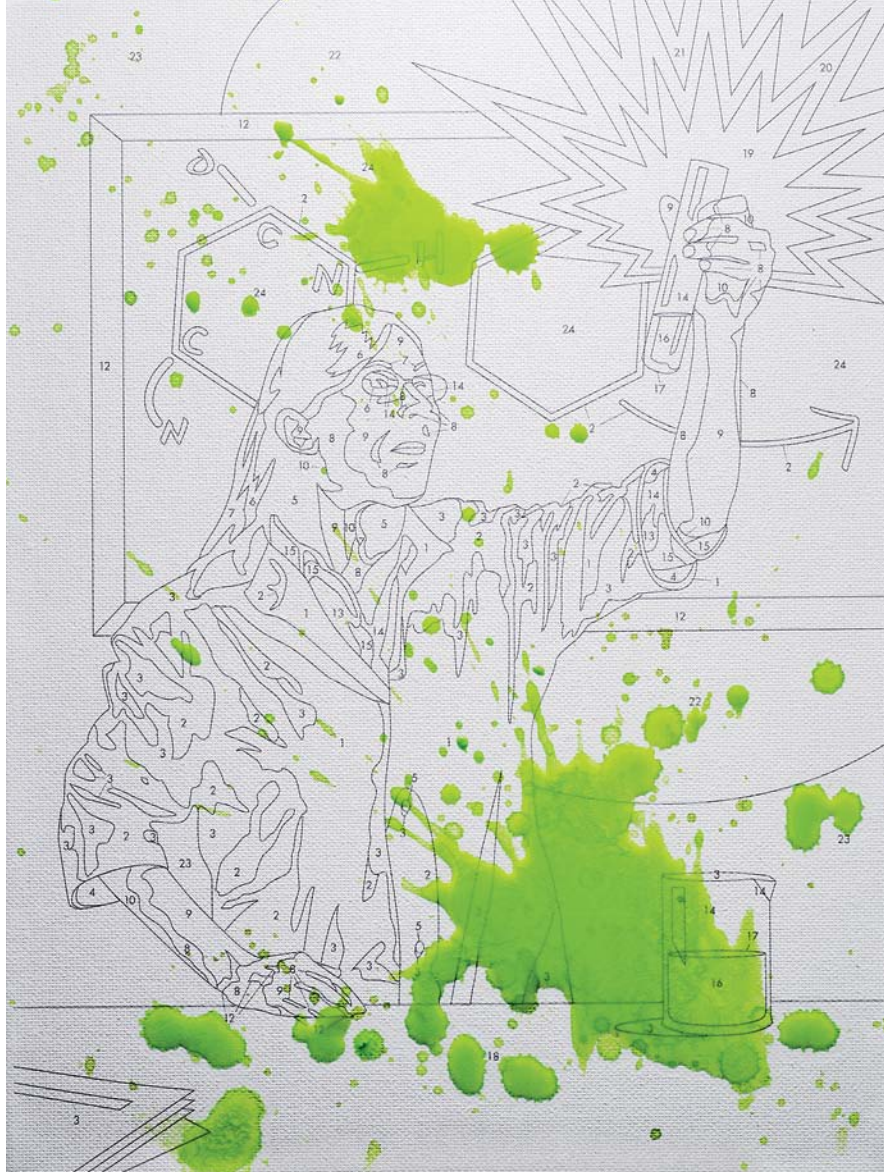
“Scientists have weak incentives to confirm the validity of old ideas. Innovative findings produce rewards of publication, employment and tenure. Replicated findings produce a shrug.”

In fiscal year 2012, the NIH’s reported annual research funding of \$31 billion was down by about 17% (adjusted for inflation) from its high in 2003. The number of applicants for NIH grants has soared almost threefold, and the NIH is able to fund fewer than one in five grant proposals. New Ph.D.s must compete for both research dollars and tenure, while senior researchers worry about being able to do the work necessary to extend their careers. Meanwhile, there may be inadequate training for the postdoctoral students who often play key research roles.

And while outright fraud may be rare, it appears to be on the increase. As a percentage of all scientific articles published from January 1973 through May 2012, retractions for fraud or suspected fraud increased tenfold, according to a study Fang and his colleagues published in *Proceedings of the National Academy of Sciences* in October 2012. “Overt dishonesty is the extreme,” Fang says. “The broad problems of reproducibility have more to do with how the work is presented and how rigorously it has been obtained because of time pressures and the importance of getting positive results.”

**C**omplicating debates about the reasons for low rates of replication is uncertainty about the magnitude of the problem. In a 2005 essay in *Public Library of Science (PLOS) Medicine*, John Ioannidis, an epidemiologist and professor at Stanford School of Medicine, argued that most published research findings are false, and used statistical models to underscore issues with how studies are conceived and designed. In 2009, Ioannidis and colleagues zeroed in on the repeatability of 18 studies of gene expression published in *Nature Genetics* in 2005 and 2006. Insufficient data made replication impossible for 16 of the papers.

In 2011, German pharmaceutical company Bayer HealthCare reported in the journal *Nature Reviews* that its scientists had been unable to reproduce nearly three-quarters of 67 published studies in cardiovascular disease, cancer and women’s health. In most cases, the inability to replicate results led to the termination of research efforts, a trend that may help explain why success rates for clinical drug trials have been declining. “Bayer HealthCare has become more cautious when working with published research targets,” says Khusru Asadullah, head of global biomarkers at Bayer’s Berlin headquarters and



New research by Ellis and a team at MD Anderson published in *PLOS ONE* in 2013 provided yet another perspective on the reproducibility problem. They reported that half of more than 400 respondents at the institution said they had been unable to replicate at least one published study. Seventy-eight percent of the scientists had attempted to contact the authors of the original scientific paper to identify the problem, but only one-third received a helpful response. More than 40% reported difficulties finding an outlet to publish findings that contradicted previous results. Such problems increase the likelihood that “suspect findings may lead to the development of entire drug development or biomarker programs that are doomed to fail,” the authors wrote.

One of the biggest problems, according to researchers at Oregon Health & Science University, is a lack of basic instructions for duplicating experiments. Their study, published in the journal *PeerJ* in 2013, examined the methods sections of several hundred articles from more than 80 journals

and found that almost half of the articles fell short in identifying all of the materials used. They also noted that methods sections had no standard guidelines and varied from one journal to the next, and were often affected by space limitations.

Ellis notes another hurdle to replication: the failure to include negative data in papers. Journals don't like to publish flawed data, but knowing an experiment sometimes failed, and why, could help other researchers when they run into trouble.

an author of the *Nature Reviews* article. “Targets now have to be better validated internally before we start big projects.”

Last year, Lee Ellis of MD Anderson and C. Glenn Begley, former head of global cancer research at pharmaceutical company Amgen, chronicled in the journal *Nature* how Amgen scientists attempted to replicate 53 landmark cancer studies and found that they could confirm only six. The scientists even consulted with the original investigators, who in some cases were unable to repeat their own experiments. But because Amgen investigators were bound by confidentiality agreements, the paper left many unanswered questions. “They didn't reveal a list of which studies they couldn't reproduce,” Fang says.

Begley, now chief scientific officer at TetraLogic Pharmaceuticals, has since provided more details, and he published his analysis, “Six Red Flags for Suspect Work,” in *Nature* in 2013. “If researchers got the results they liked in the first experiment, they usually didn't repeat it,” Begley says. Much of today's research isn't fudged, he says, or fraudulent: “It's lazy and sloppy.”

Several prominent journals, including *Nature*, *Science* and *Science Translational Medicine*, are now adopting guidelines to ensure the disclosure of all technical and statistical information that is crucial for reproducibility. *Nature* now provides more space for methods information and requires more precise information from authors. And to publish in *Science*, senior authors must sign off on a paper's primary conclusions. The peer review process is also being scrutinized, with the aim of “increasing transparency,”

## When *Science* published technical comments disproving part of an Alzheimer's study, they sparked publicity that might encourage other replication work.

particularly in analyzing researchers' statistical measures, says Meagan Phelan, a spokesperson for the American Association for the Advancement of Science, which publishes *Science*.

Meanwhile, the Reproducibility Initiative has received \$1.3 million in funding from the Laura and John Arnold Foundation to replicate key findings from 50 landmark cancer biology studies. The foundation is also financing a related effort, the Reproducibility Project, which Brian Nosek helped establish, that is bringing together more than 180 academic psychologists through a network called the Center for Open Science to replicate 100 papers published in three prominent journals.

Some scientists have expressed reservations about such efforts, citing potential conflicts of interest that could arise from a private company acting as gatekeeper. Others suggest getting funding agencies such as the NIH to support development of technology to make biomedical research data electronically available.

Indeed, reproducibility is a high priority at the NIH, says Lawrence Tabak, the agency's principal deputy director. Some NIH institutes are looking for ways to improve peer review processes for grant applications and to provide better training in research methods for scientists. Tabak also says the agency is considering how it could support the validation of preclinical studies linked to proposals for large, expensive clinical trials.

Medical institutions could also help reform the replication process, suggests Bruce Booth of Atlas Venture. Technology transfer offices, which universities have set up to support researchers in patenting their work and creating private companies, might redirect some of their resources to research

replication, Booth says. "If they could show third-party data supporting a lab's findings, the prospects for funding would increase significantly, and failure rates could fall," he says.

Last May, when *Science* published the technical comments disproving part of the Alzheimer's study, they sparked publicity that might encourage other researchers to undertake the often thankless task of attempting replication. But the same issue also included a response from Gary Landreth, a neuroscientist at Case Western Reserve School of Medicine and lead author of the original study, who speculated that the replication failures might be related to how researchers prepared and administered the drug. Meanwhile, other groups are still trying to replicate the study's results, according to Landreth, who says that findings presented at recent conferences have confirmed bexarotene's impact on memory in mice. The original research has also spawned investigations into whether the drug might be helpful in treating other diseases.

Nor has the controversy surrounding the lab's original findings deterred investigations into whether bexarotene could help human Alzheimer's patients. In one small clinical trial, Landreth and his lab are looking at the drug's effect on the brains of healthy human volunteers, while the Cleveland Clinic Lou Ruvo Center for Brain Health in Las Vegas is recruiting patients with moderate Alzheimer's for another trial. These lines of inquiry would ideally be based on much more than that one tantalizing result. Yet pursuing them while replication efforts continue is better than nothing, many would argue—and even if researchers are met with failure, failure still counts as a result. ■

### → DOSSIER

1. "Why Most Published Research Findings Are False," by John Ioannidis, *PLoS Medicine*, August 2005. In this seminal study on replication, Ioannidis, a Stanford University epidemiologist, uses statistical models and other key factors to demonstrate why the vast majority of published research findings are false.
2. "Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets?" by F. Prinz, T. Schlange and K. Asadullah, *Nature Reviews Drug Discovery*, September 2011. In an industry analysis, three Bayer HealthCare scientists report that in-house experiments over four years failed to replicate two-thirds of 67 research studies in the fields of oncology, women's health and cardiovascular diseases.
3. "Drug Development: Raise Standards for Preclinical Cancer Research," by C.G. Begley and L. Ellis, *Nature*, March 29, 2012. A chronicle of scientists at American drug company Amgen who tried to replicate 53 studies they considered landmarks in the basic science of cancer—and were able to replicate only six.