THE CHRONICLE OF HIGHER EDUCATION
Research

March 16, 2015

Amid a Sea of False Findings, the NIH Tries Reform

By Paul Voosen

How do you change an entire scientific culture?

It may sound grandiose, but that is the loaded question now facing the National Institutes of Health, the federal agency that oversees and finances U.S. biomedical research.

While the public remains relatively unaware of the problem, it is now a truism in the scientific establishment that many preclinical biomedical studies, when subjected to additional scrutiny, turn out to be false. Many researchers believe that if scientists set out to reproduce preclinical work published over the past decade, a majority would fail. This, in short, is the reproducibility crisis.

The NIH, if it was at first reluctant to consider the problem, is now taking it seriously. Just over a year ago, the agency's director, Francis S. Collins, and his chief deputy, Lawrence A. Tabak, announced actions the agency would take to improve the research it finances. Science needs to get its house in order, Dr. Collins said in a recent interview with *The Chronicle*.

"We can't afford to waste resources and produce nonreproducible conclusions if there's a better way," he said.

The problem may be most visible in preclinical biomedicine and fields like psychology, but it's by no means limited to them, Dr. Collins added. "No matter what your science is, there's a risk here. If experimental methods are not well detailed and well planned,

then stuff can happen."

Science today is riven with perverse incentives:

- Researchers judge one another not by the quality of their science — who has time to read all that? — but by the pedigree of their journal publications.
- High-profile journals pursue flashy results, many of which won't pan out on further scrutiny.
- Universities reward researchers on those publication records.
- Financing agencies, reliant on peer review, direct their grant money back toward those same winners.
- Graduate students, dependent on their advisers and neglected by their universities, receive minimal, ad hoc training on proper experimental design, believing the system of rewards is how it always has been and how it always will be.

But many senior researchers will oppose reform, said Arturo Casadevall, a professor at the Albert Einstein College of Medicine at Yeshiva University. "Scientists themselves are playing this game because once they succeed, the rewards are so great they basically force everyone to do it."

The tight financing of the past decade has only made the problem worse, said Ferric C. Fang, a microbiology professor at the University of Washington who, like Dr. Casadevall, has called for changes in the system. "We're ending up with an ecosystem that's increasingly populated by predators," he said.

Some researchers fear that "skeptics" who dispute valid scientific work on a host of hot-button issues will seize on the reproducibility crisis as a political weapon.

"This is a very serious problem," said C. Glenn Begley, chief scientific officer of TetraLogic Pharmaceuticals, who, several years ago, published a prominent paper on how his former company, Amgen, had found it impossible to replicate more than 90 percent of potential cancer drugs.

"It really threatens the general feeling among the community in terms of how important science is," he said.

The NIH's response is wide- ranging. Its institutes are revising how they review grants, requiring far more data on experimental design, including validation of past findings that studies purport to build upon. The agency is financing new programs to teach first-year graduate students how to design experiments. It is pressing journals to raise their review standards. It is improving how researchers identify lab materials, and requiring balanced use of males and females in animal studies. And it is experimenting with different ways of financing research.

Those are all worthy steps, Dr. Casadevall said, and he applauds how quickly the NIH, once it understood the scale of the problem, has responded. But even the NIH is limited in how it can change science. "My concern is that the problem here is cultural," he said, "and the [proposed] fixes are procedural."

An Inside Agent

If the NIH could be said to have an internal whistle-blower on the problems with preclinical research, Shai D. Silberberg would be a candidate for the role. A decade ago, Mr. Silberberg, a biophysicist, became a program director at the National Institute of Neurological Disorders and Stroke, and he was shocked to see what was considered high-quality science, especially in studies using animals. Many experiments supporting proposed clinical investigations used low sample sizes and were prone to bias.

As it happened, neurology researchers were ahead of the curve in pointing out those problems. Scientists in Australia and Scotland had begun documenting flaws in animal studies that had seemed to support promising, but failed, stroke treatments. A similar effort found parallel problems in spinal-cord research. And ALS researchers, seeking a cure for Lou Gehrig's disease, went back and reproduced studies on more than 70 promising drugs. They

found no real effects.

"Zero of those were replicable," Dr. Collins said. "Zero. And a couple of them had already moved into human clinical trials, which is like — oh, man. You've started maybe even putting people at risk for exposure to a compound for which there is really no compelling data that it was going to help. This can't be."

Armed with such data, Mr. Silberberg made a case for change. He won over the institute's director, Story C. Landis, who put her grant reviewers on notice that higher standards were expected. But it was clear to Ms. Landis that the problem wasn't limited to neurology; all of NIH needed to change. That culminated in a 2012 meeting with editors from journals like *Nature, Science,* and *Cell;* financing agencies; and prominent reviewers. The first day, everyone blamed each other. The second day, they started talking solutions.

Though replication has now become an NIH-wide concern, Mr. Silberberg remains enmeshed in the agency's efforts to improve transparency in detailing how studies are conducted. He gave 24 talks about the effort last year.

"I view this like being a Johnny Appleseed," Mr. Silberberg said. "You've got to plant the seeds and let them grow."

Training Young Scientists

The NIH has realized it has to get to scientists while they're young. To that end, the agency has asked researchers to develop training modules for young scientists on proper experimental techniques, and will soon choose 25 of them to finance in the hope of creating a library of material, said Jon R. Lorsch, director of the NIH's National Institute of General Medical Sciences.

The agency will also offer grants this year for universities willing to make more-radical shifts, such as changing how they teach first- and second-year graduate students to emphasize proper use of

statistics and experimental design.

Indeed, more than any part of the scientific system, the universities have been ignoring the replication crisis, Dr. Begley said. "I don't see the institutions taking this on seriously."

Currently, the experimental rigor of a graduate student's education is up to the lab where he or she resides, and few universities teach experimental design. It's an apprenticeship, yet often the master of the shop is sequestered, filing grant applications. Dr. Begley predicted that the first university to tackle the problem would draw much interest from venture capitalists burned by patents that petered out. So far, that incentive hasn't been enough.

"As long as universities think that the way for investigators to get money is to publish in *Nature* and *Science,* then that's what they're going to look for in investigators," Mr. Silberberg said. "They want that money."

Dr. Begley would like to see the NIH go further, requiring universities to ensure a minimum research standard among their employees. Even if that step cut in half the amount of research universities produced, if it increased the science's quality by a commensurate amount, Dr. Begley would take it.

The NIH has also pushed scientific journals to raise their review standards for life-science studies. More than 100 journals have agreed to elevated principles, many of which stem from the neurology institute's recommendations. Most of the journals are instituting checklists that will require submissions to run down and justify the decisions made in designing the study.

Checklists can prevent all kinds of human-scale problems, Dr. Casadevall said. "The next time you take a flight, look left as you're boarding the plane. The pilots are doing a checklist."

The institutes are also exploring better ways of identifying materials used in the lab: cell lines, antibodies, specialty chemicals. By some estimates, up to 30 percent of cell lines are mistakenly labeled.

"They get passed around from lab to lab," Dr. Collins said. "There's a little mix-up in someone's incubator, and then, 10 years later, somebody's been working on this cell line and it turns out it isn't what they thought it was at all."

The NIH is considering a focused investment in technology that would make it quicker and cheaper to check the validity of a cell.

While it makes those tweaks and more, the NIH has also been experimenting with different ways of financing research. Knowing that the unstable environment can drive researchers to cut corners or oversell work, the agency has tried systems that finance individual investigators, rather than projects. That may allow more time for careful science, not grantsmanship.

"In some ways, ultimately, that may be the most important thing we can do," said Mr. Lorsch.

But perhaps the agency's most radical step is buried in its education proposals. Not only does it want training on experimental methods; it wants lessons on the sociology of science. Why do promotion and tenure work like they do? Are there alternatives? Seeds must be planted. Students need to know that it doesn't have to be this way.

"We want them questioning, wanting to fix what's broken in the system," Mr. Lorsch said. "If they don't hear what the problems are, and don't hear their professors reflecting on them, it's going to be much harder to change things."

Paul Voosen is a senior reporter covering the sciences. Write him at paul.voosen@chronicle.com; follow him on Twitter @voooos; or see

past work at voosen.me.

21 Comments The Chronicle of Higher Education Login

Recommend Share Sort by Oldest

Join the discussion…

schultzjc · 5 days ago

This is a terrible piece of "journalism". The writer appears to have done little homework. For example, I would not regard any of the 'perverse incentives" he lists as generally true, and I think most researchers would agree:

Researchers judge one another not by the quality of their science — who has time to read all that? — but by the pedigree of their journal publications.
High-profile journals pursue flashy results, many of which won't pan out on further scrutiny.
Universities reward researchers on those publication records.
Financing agencies, reliant on peer review, direct their grant money back toward those same winners.
Graduate students, dependent on their advisers and neglected by their universities, receive minimal, ad hoc training on proper experimental design, believing the system of rewards is how it always has been and how it always will be.
The bullet that comes closest to truth is how universities reward faculty. The rest - bah. I can't imagine writing a piece like this without talking with MANY researchers, successful and not so. If that were done, I'm pretty sure the

see more

1 · Reply · Share ›

sciencegrad ↗ schultzjc · 4 days ago

I agree about the stats, but I think it's due to benign lack of skills. In my experience as a grad student, none of the professors I collaborated with would use correct stats because it would make their results nonsignificant. They said all that matters is getting into a top journal at any cost and that if I want a faculty job, I should be quiet and do what they say.

1 · Reply · Share ›

Hyptiotes ↗ sciencegrad · 4 days ago

I think cell and molecular biology is particularly remiss in acknowledging variation in data and the necessary statistical analysis and experimental design that goes with that. Running an experiment three times and calling it true when it "works" two of the three times, does not constitute a valid method. I've run into many cell and molecular biologists that, when I ask them why they run three samples, they say, "because it gives you a standard deviation". Ug. Of course, if you look at the sample size necessary to generate sufficient statistical power to draw meaningful inferences, it is often outside the financial scope and time horizon for most researchers. As a result, we (the collective we) publish dodgy pilot data and call it a real study.

2 · Reply · Share ›

Jack C. Schultz ↗ Hyptiotes · 3 days ago
Yes!!! I've known investigators to rerun experiments until they get the "right" result!
· Reply · Share ›

12105442 · 4 days ago
I liked the quote, "...Some researchers fear that "skeptics" who dispute valid scientific work on a host of hot-button issues will seize on the reproducibility crisis as a political weapon..." But how do we know what's is valid scientific work and what isn't if reproducibility of results isn't carefully evaluated? Research work on the "right" side of a host of hot-button issues may have gotten a pass. Thus the "skeptics" may have a point.
1 · Reply · Share ›

Hyptiotes ↗ 12105442 · 4 days ago
They need to start giving out grants for replication and they must always be given to a different lab group. Another cheaper alternative would be to have an entire category of science publications that are based on picking apart other published work. In other words, publish the reviews, not just the original paper.
· Reply · Share ›

silverquille37 ↗ Hyptiotes · 3 days ago
Indeed! As long as "innovation" remains an important criterion for NIH grant funding, the paucity of replication studies will remain.
(See: http://grants.nih.gov/grants/p...
· Reply · Share ›

11889694 · 4 days ago
From what I've seen almost every university teaches experimental design and requires every science graduate student to take the course(s). On the other hand the typical (but not every) M.D. has had virtually no training in how to be a scientist, so maybe that's what the author was thinking about. If reproducibility is the problem, and that appears to be the case, then an obvious solution is to give scientists the incentive to repeat published experiments. No culture change is needed at the faculty level. Provide the funding for such work and academic researchers will immediately do it.
· Reply · Share ›

digithead99163 ↗ 11889694 · 4 days ago
Taking a course in experimental design does not mean you can perform experimental design any more than taking wood shop in high school makes one a carpenter. We need to do more in the apprenticeship of young scientists.regardless of discipline in how to set up proper design of experiments from beginning to end. Moreover, that null findings are just as good as statistically significant findings.
1 · Reply · Share ›

11889694 ↗ digithead99163 · 4 days ago
That was a response to Mr. Voosen's statement "few universities teach experimental design.", i.e. a comment on the quality of the reporting.
· Reply · Share ›

procrustes · 4 days ago

From the tenor of comments so far, one can see the truth hurts. There is too much corner-cutting, and that is well before addressing the widespread failure of peer review (visited Retraction Watch lately?), and too many instances of outright fraud. If science doesn't get its house in order, there is no reason to keep pouring tax dollars into it.

2 · Reply · Share ›

Bill ↗ procrustes · 4 days ago

"If science doesn't get its house in order, there is no reason to keep pouring tax dollars into it."

Or, making hugely expensive public policy decisions based on it.

3 · Reply · Share ›

Helen · 4 days ago

Near the end of my days as a practicing scientist, I had two experiences with manuscript review that stood out as contributing to this problem. The information is anecdotal, to be sure, but I wonder how common such experiences are.

First, I was asked to review a manuscript for a lower tier journal. I rejected the manuscript as written due to two glaring problems. One was a conclusion drawn from comparing, more or less, apples and oranges (the baseline was determined using one assay/output, and the relevant condition was tested using a different assay/output!!), and the second was the main claim that two things were different was based on a "difference" that was not statistically significant. I told the editor that I was willing to review a revised manuscript that addressed these issues. A month or so later, I received the "revised" manuscript that added data from an additional experiment, but neither changed the errors I had pointed out nor addressed my comments in a written argument. I rejected it again. Later, in an off-line conversation, the editor (a friend of mine) said he published the paper anyway because he found someone else to review the paper who wasn't concerned about the errors because "it's just (insert name of lower-teir journal) and the impact factor is

see more

1 · Reply · Share ›

phonenear ↗ Helen · 4 days ago

Lots of us as practicing scientists have analogous anecdotes, or worse. One of mine hinged on "representative data" (versus quantitation buried in some supplemental figure (maybe #17 out of >20) in a very prestigious journal (named in the article above), and the editor handling the manuscript just let it go (let the author argue it away over the phone!). A current one involves auditing an alleged P value ($p<0.05$) at another high-profile journal (one that even mixes clinical and basic science efforts) and finding that for the means and variances presented, there was a serious problem (along with others) - and yet it is conditionally accepted. Sadly enough, I wasn't even surprised. I wish I could agree with schultzjc (above), and many NIH "fixes" are worse than the disease itself, but fundamentally I have to agree that a culture change is needed - not just with the reviewers, but also with editors in how they handle basic issues such as those you've mentioned, Helen, and concerns raised about how scrupulously authors are presenting fair facts and honest stats.

· Reply · Share ›

Socratease2 · 4 days ago

Just in case the "climate change denier" or creationist camps are chomping at the bit to dismiss the whole of scientific research across all disciplines and throughout history, this article is speaking specifically to issues concerning the reproducibility of bio-medical research. But individual study bias and error occur constantly so that is why no one study is going to be considered foundational or paradigm shifting no matter what the conclusion. No study is perfect either and all researchers, especially those with questionable methodologies, testing and analysis, should include a "mea culpa" section saying why the findings are limited, what future studies should be done to further explore the evidence, etc. That shoddy studies get published is not the problem, no one with any credibility in the field is taking them at face value anyway. At least they shouldn't be. The problem would seem to be that irrelevant and distracting studies clog the filters of science communication and they create a lot of work to decipher if there is actually something worth trying to reproduce from the growing volume of published studies.

2 · Reply · Share ›

Hyptiotes ↗ Socratease2 · 4 days ago

Science as a process is inefficient, expensive, and often ends with ambiguous results. Despite this, it does a pretty good job compared to the alternatives--blood-letting, incantations, and tea-leaf readings.

1 · Reply · Share ›

Socratease2 ↗ Hyptiotes · 4 days ago

Not to mention listening to gaseous oracles, reading the patterns of fractures in the burned scapulas (scapulamancy) of sheep, goats or other animals or, worst of all, spending a day at Ken Ham's "Humans walked with dinosaurs" theme park in Kentucky.

1 · Reply · Share ›

phonenear ↗ Socratease2 · 4 days ago

Unfortunately, reading review articles and seeing the way some work with real shortcomings in my own area is uncritically converted into stated Truth, I have to say that your statement "no one with any credibility in the field is taking them at face value anyway" is wishful thinking, as revealed by the next sentence (you're right - they shouldn't be taking them at face value but they are).

Reply · Share ›

Socratease2 ↗ phonenear · 3 days ago

Well, now I am worried!

Reply · Share ›

igopogo · 3 days ago

On path to correction of such noxious practices would be in the preliminary examination phase of the Ph.D. I am aware of a few departments in a few places where the candidate must take a published article (with its data set on file with the journal) and attempt to replicate its result. The written paper that results is itself graded. The knowledge that someone somewhere might attempt to replicate one's own future paper might be salutary.

Reply · Share ›

wskocpol ↗ igopogo · 3 days ago

This would be just as effective as auditing 1% of tax returns with no follow-up enforcement. There are lots more papers than graduate

follow-up enforcement. There are lots more papers than graduate students taking a prelim, and attempting to replicate a result is different from taking an unbiased look at what interesting results emerge from, or are refuted by, the data. If the paper and the replication agree, should we reward or punish the student?