In recent years, there has been an upswing in both retractions of scientific papers and failures to reproduce published results. Ruth Williams asks researchers what is going on and what can be done.

Back in the days when I was a bench scientist, I remember feeling happiest not when I had made a new discovery but when someone else repeated it. Until that point, I would have creeping doubts that perhaps I had done something wrong or missed something important. When someone independently reproduced my results, it meant I could relax.

I am perhaps a little prone to worry, but luckily, my anxieties had a scientific basis. Like my colleagues, I knew that any eureka moment would be just the start of a potential scientific story and that the plot would go nowhere until the initial results were confirmed.

Reproducibility in science is as important as any new hypothesis or discovery. Indeed, the three are inextricably linked like spokes on a wheel propelling science forward, and it is reproducibility that prevents that wheel from meandering down irrelevant paths.

It is disquieting then that the results presented in a large number of scientific papers seem to be irreproducible. In the August 28th issue of Science, a report by the Open Science Collaboration—a contingent of the nonprofit Center for Open Science—revealed that of 100 published psychology experiments the group attempted to replicate, just one third to one half were successfully reproduced. Similarly, attempts by pharmaceutical and biotechnology companies to confirm published preclinical data have failed more often than not. Biotechnology company Amgen, for example, succeeded in replicating the results of just 6 of 53 (11%) selected high-impact papers, where pharmaceutical giant Bayer HealthCare was only able to reproduce findings of ≈25% of research papers.

Furthermore, these failed replication attempts followed a decade in which retractions of scientific papers had reportedly shot up 10-fold.

So, why is irreproducibility so common? And what, if anything, can journal editors and scientists do to improve matters?

Tricky Techniques

“There are many reasons,” says Aruni Bhatnagar, Professor of Medicine at the University of Louisville in Kentucky and Deputy Editor of Circulation Research. Among other things, he says, it might be due to a “lack of detail in published procedures, the unavailability of exactly the same reagents, inadvertent modifications to the original protocol and sometimes a lack of expertise.” Often it simply comes down to the fact that some techniques are really difficult, he says. “The methodology for accomplishing complicated experiments is usually very complex.”

Even small changes to a protocol may hinder its reproduction, says Linda Demer, Professor of Cardiology, Physiology and Bioengineering at the University of California, Los Angeles. As examples she lists, “different cell types, different tissue culture plastic, different reagents, different serum supplements to tissue culture, different backgrounds of engineered mice, different Chow, and different housing arrangements (such as light cycle schedules and bedding).” With regard to mice, differences in the age of the animals or the administration route of drugs can also affect outcomes, says Hori Masatsugu, Professor-Emeritus at Osaka University and Invited Professor at Osaka University of Pharmaceutical Sciences in Japan.

Such subtle differences to a technique may be to blame for the difficulties in reproducing the results of a 2010 Cell paper that found the direct conversion of mouse fibroblasts into cardiomyocytes by the addition of three transcription factors. In a 2012 Circulation Research paper, Sean Wu of the Harvard Stem Cell Institute and colleagues revealed that they were unable to reproduce the method, while another group claimed it could perform the technique, and a third group reported only partial success.

Similarly, irisin, identified in 2012 as an exercise-induced hormone that burns fat, and GDF11, identified in 2013 as a rejuvenating factor, have both been plagued with inconsistent follow-up studies and doubts over experimental techniques.

These discrepancies could simply be teething pains of new techniques, but how could such problems with difficult methods be resolved? “We could start by encouraging authors to publish their methods in as much detail as possible,” suggests Bhatnagar. Details as specific as the source and even lot numbers of reagents should be included, he says. And there is no excuse not to, says Alan Tall, the Tilden Weger Bieler Professor of Medicine at Columbia University in New York. “With online supplements there is no reason that extensive methodology cannot be published.”

DOI: 10.1161/CIRCRESAHA.115.307532
“There is almost a sacred obligation to clearly explain our technical details in the Methods or Procedures sections of our papers,” agrees Irwin Gelman, Chair of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, New York, writing in The Scientist magazine. “Without a doubt, there has been a steady erosion of this process, making it difficult, if not impossible to recapitulate the findings of others,” he writes.

Organizations, such as Center for Open Science, are endeavoring to improve transparency in methodology by, among other things, encouraging journals to adopt their recently devised Transparency and Openness Promotion guidelines. But, another solution, says Demer, is that “editors could possibly generate position papers in which leaders in a field agree on a particular protocol or approach as the standard.”

Demer also points out, however, that too much standardizing may not be a good thing. “If we all used the same technique, then we might have the impression that a finding is strong, when it depends entirely on a specific condition,” she says. “In some respects, these differences are helpful because, if a finding turns out to be reproducible across labs despite differences in models, it is evidence that the finding is robust.”

Although Demer’s point is certainly true, it does not necessarily follow that a finding is not consistent across different models is not robust. That is to say, a subtle variation in conditions that results in a failure to reproduce results may be important in itself. It might even lead to a new discovery or, as in the case of Robert Furchgott, to a Nobel Prize, says William Chilian, Chair of Integrative Medical Sciences at Northeast Ohio Medical University. “Furchgott’s new technicican could not repeat previous work (performed) in his lab,” explains Chilian, “and rather than being superficial, Dr Furchgott got to the bottom of the problem and found endothelium-derived relaxing factor—that we now know is nitric oxide.” The discovery led to Furchgott winning the Nobel Prize.

Chilian thus absolutely supports Demer’s position paper idea. “We need standard practices for methods as a way to make sure that work is repeatable,” he says. “My prediction is that a lot more work would be repeatable if labs used the same techniques and models.” At Circulation Research, Roberto Bolli, Chilian, and the other editors have, in fact, endeavored to implement such a publication policy, which “if successful will be an official position of the American Heart Association,” says Chilian.

Dodgy Design

Besides methods being difficult to follow or master, another reason for irreproducibility can be that the original experiments themselves were poorly designed. For example, “sample sizes may be insufficient in the original experiment and thus apparent differences are not real and cannot be reproduced,” says Tall. Jeffrey Robbins, Executive Co-Director of The Heart Institute at Cincinnati Children’s Hospital, Ohio, adds that for some experiments, there may be insufficient use of blinded observation.

Such deficiencies might stem from “inadequate supervision of the young scientists and trainees by the senior mentors… [or] inadequate training of the young investigators, and sometimes even the senior investigators, for proper experimental design and robust methodology,” suggests Ali Marian, Director of the Center for Cardiovascular Genetic Research at the University of Texas Health Sciences Center in Houston. Chilian also highlights general “carelessness” as a cause.

And the editors of Nature would agree. In a 2012 editorial, they complained of an increase in the incidence of sloppy mistakes in submitted papers, citing, “Incorrect controls…and improper use of statistics”—such as, “the failure to understand the difference between technical replicates and independent experiments.” By extension, they say, these sloppy papers “reflect unacceptable shoddiness in laboratories.”

It is possible that such shoddiness is a result of scientists rushing to publish their data. For example, Marian blames a “dictum of publish or perish,” suggesting that there is an “excessive emphasis by peers and organizations on publication productivity [which] leads to generating data just for the sake of publication rather than discovery.” All too often, he says, “research is used to advance one’s career” instead of being based on “a sincere belief in scientific discoveries.” Worse still, he adds, “often trainees [and] postdocs are set to prove the mentor’s hypothesis and not to test it.”

Whatever the cause of a poorly designed experiment—be it insufficient training, shoddiness, or even an innocent oversight, the result can be that large amounts of time and money are spent not only by those performing the initial experiments, but by researchers trying to repeat the work. Indeed, a PLoS Biology paper published this June estimated that $28 billion per year is spent on irreproducible research in the United States alone.

To combat poorly designed experiments, journal editors and reviewers could be more stringent, says Tall, suggesting that, for example, “the issue of sample size can be more commonly scrutinized.” Ultimately, however, the responsibility lies with the researchers, says Marian. “We need to de-emphasize publish or perish [and] train ourselves, and our trainees, in robust experimental design and data interpretation,” he says.

Bad Behavior

Although poor methodology is shameful in itself, it is certainly a far cry from deliberate data manipulation and fabrication, which brings us to our final reason for irreproducible results.

There are few people, let alone scientists, who have not heard of Woo Suk Hwang, the former professor at Seoul National University who became infamous for fabricating data that suggested his laboratory had created human embryonic stem cells from cloned human embryos. Many will have also followed last year’s saga of the two Nature papers describing stimulus-triggered acquisition of pluripotency. The papers were rapidly retracted after their lead author, Haruko Obokata, was found guilty of misconduct. And shortly thereafter, in the wake of the high-profile investigations, her coauthor and supervisor Yoshiki Sasai committed suicide.
Such cases make headline news. But should we really be worried about fraud? Robbins says, no. “People continue to moan and gnash their teeth… [and] the reporting has become more breathless as this happens in high-publicity journals,” he says, “[but] I don’t think that fraud has become a serious problem—at least no more serious than it has been.”

While fabricating data is probably “relatively rare,” says Bhatnagar, perhaps more common is manipulation of results, says Tall, such as “the selective removal of some aspect of the data.”

To catch such misconduct, Bhatnagar suggests that “the journal could ask for original gels etc that could be electronically validated to detect duplication or tampering.” Stefanie Dimmer, Director of the Institute of Cardiovascular Regeneration, in Frankfurt, Germany, adds, “This should be simply normal routine and would help to avoid mistakes.” Some publications, such as the Journal of Clinical Investigation, the Journal of Cell Biology, and the Journal of Experimental Medicine, have already adopted such policies. However, Bhatnagar also stresses, “ultimately, you have to trust the authors. It is their work and their reputation on the line.”

He believes that over policing could even be counterproductive. “Excessive policing is likely to slow down the review process and place greater burden on the authors and in the end might lead only to more clever evaders.” Marian agrees: “Overzealous editors do more harm than good,” he says. “It is silly to think that you can catch various technical details of the complicated experiments. Those who catch the obvious ones might get the wrong impression that they are vigilant and falsely praise themselves.”

Marian also points out that “some of us refuse to submit manuscripts to journals that do not trust us with our data.” Because, he says, “it is an insult to authors integrity [and it] poisons the ambience between the journals and the authors.”

Katrin Deinhardt, a neuroscientist at the University of Southampton in the United Kingdom, has firsthand experience of such a poisoned ambience. When working as a PhD student at Cancer Research, United Kingdom, she submitted an article to the Journal of Cell Biology, and was contacted by administrative staff who claimed that one of her photos had some “irregular background.” “I did not mind having to send in the original data,” says Deinhardt, who had not manipulated the image. “What I did mind was that it [the letter] was written in a tone basically accusing us of fraud when there was a very simple explanation.”

It is possible that high-profile cases of scientific misconduct have made certain editors overly skeptical. Perhaps they worry that a fraudulent paper will end up on their pages and harsh questions will be asked. But the skepticism and worrying should stop says, Marian. “The principle governing our society, whether scientific or otherwise, must be trust,” he says.

There are those outside the scientific community, however, who argue that some form of external regulation to prevent scientific misconduct is necessary—in much the same way that there are external regulators overseeing health and safety. Indeed, members of the House of Commons Science and Technology Committee in the United Kingdom have called on the government to implement such an external regulatory body.20

Some scientists balk at this idea. “I do not like the idea of an agency controlling the data,” says Dimmeler. “We already spend too many dollars and euros on bureaucrats watching our experiments—for issues of animal ethics, the control of genetically modified organisms, etc. This would create only more bureaucracy, but does not help at all.” Marian agrees, adding, “More regulation of any shape or form is likely to be damaging, as it means less time for research and less time for thinking and creativity.”

Whether such external regulation ever comes into effect, Marian and the other scientists interviewed for this article certainly think that scientific misconduct is the result of a few bad apples. By far, the main cause of irreproducible data, they all agree, lies in the design and the details of the experiments. The good news is that resources, such as the Open Science Framework—a free cloud-based project management solution for researchers from the Center for Open Science—and the implementation of The Center’s Transparency and Openness Promotion guidelines by journals, are starting to make the sharing of experimental details both easier and unavoidable. And that’s important because the more open and transparent research is—whether by virtue of guidelines or firm rules—the more likely the wheel of science will maintain a true course.

Disclosures

None.

References


Can't Get No Reproduction: Leading Researchers Discuss the Problem of Irreproducible Results
Ruth Williams

doi: 10.1161/CIRCRESAHA.115.307532

_Circulation Research_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2015 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/117/8/667

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation Research_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation Research_ is online at:
http://circres.ahajournals.org/subscriptions/