

Repairing reproducibility

By C. Simone Fishburn, Senior Editor

Although the NIH is taking preliminary steps to improve the reliability of preclinical data, many researchers are not waiting for reproducibility guidelines and are taking up the mantle to raise standards at their own institutions.

Francis Collins and Lawrence Tabak recently outlined a series of measures they are planning to implement at the NIH to combat poor reproducibility in preclinical research.¹ The crux of the problem, they said, appears to be rooted in practices, policies and attitudes at academic centers, scientific publishers and funding agencies.

Collins is director and Tabak is principal deputy director of the NIH.

“Reports of challenges in reproducibility have been around for several years, but over the last couple of years they have increased in prevalence. We decided the agency needed to articulate a way forward,” Tabak told *SciBX*.

Indeed, industry researchers highlighted the frequent irreproducibility of preclinical data in two analyses published in 2011 and 2012. In the earlier article, **Bayer AG** scientists managed to reproduce published data in only about 25% of 67 projects.² The next year, scientists from **Amgen Inc.** and **The University of Texas MD Anderson Cancer Center** reported that the scientific findings of 53 landmark oncology studies were upheld in only 6 cases (11%).³

Since then, privately and publicly funded groups have launched initiatives to address the problem. The Reproducibility Initiative, a joint venture by **Science Exchange**, **PLOS** and **figshare**, was launched in 2012 to allow academic researchers to have their work replicated by a third party.⁴

The **Global Biological Standards Institute** was founded that same year to develop biological standards for basic and translational research. This year, the institute set up task forces to create standards for tools such as cell lines, antibodies and next-generation genome sequencing.⁵

Now, the NIH has decided to take a leadership position in trying to drive change, focusing first on better training for scientists and improving the rigor of the institute’s grant review teams.

Home testing

The NIH is starting by testing ideas and introducing new procedures for its own intramural scientists and grant review teams. The aim is to develop recommendations for use by other institutions.

For example, the NIH is developing a training module for postdocs, graduates and other students to teach good experimental design and record keeping. The module is focused on increasing the reproducibility and transparency of research findings and will become part of the mandatory training at the NIH. It will be made publicly available by the end of 2014.

Tabak said that the emphasis is on recommendations rather than rules, and he does not expect punitive actions for failure to comply. “We prefer to use the bully pulpit to convince people to do the right thing and make it obligatory when there’s no alternative. Some of the recommendations could become a best practice that we expect people to adhere to,” he said.

The NIH already has two types of pilot studies under way related to funding for clinical trials based on preclinical studies.

In one, the NIH is performing after-the-fact studies, Tabak told *SciBX*, in which the institute is replicating select experiments in published papers on which applications for clinical trials are based. Tabak said that the pilot program could provide an indication of what proportion of studies are reproducible and give a metric that can be followed over time to monitor improvement.

In the second type of pilot study, grant review panels will have an additional focus on evaluating the scientific premise of applications, including the statistical basis of preclinical experiments. If proposed clinical studies are based on underpowered preclinical experiments, that could affect the decision of whether to fund or not, Tabak said.

The pilot studies involve multiple NIH institutes, and the NIH plans to assess the outcomes and decide on what practices to adopt by year end.

Tabak said that he could not yet estimate what it would cost the scientific establishment in terms of time and money to replicate

selected experiments.

In addition, the NIH is aiming to increase transparency through PubMed Commons, an online discussion forum about published articles that was launched last December, and a planned Data Discovery Index to house primary data on which published manuscripts are based.

Lee Ellis told *SciBX* that the ability to comment on papers in PubMed Commons is a big step forward.

According to the Collins and Tabak paper, so far PubMed Commons has had about 2,000 people sign up and has received about 700 comments.

Ellis said, “These numbers are relatively low, but hopefully as more people become aware of this they’ll be more willing to comment on papers.” Although it will take some time to see the full effects of such comments, he said that the ability to stimulate discussion about published papers is healthy for the field.

Ellis is a professor of surgery and molecular and cellular oncology at MD Anderson and vice chair of the cancer research cooperative group **SWOG**. He was coauthor on the 2012 paper on reproducibility in oncology studies.

It takes a village

According to Tabak, the NIH initiative is not a reaction to political pressure and instead is part of a community-wide response to a problem that involves many stakeholders.

“NIH alone can’t solve this, but in partnering with journals, investigators and others we can make headway,” he said.

“Reports of challenges in reproducibility have been around for several years, but over the last couple of years they have increased in prevalence. We decided the agency needed to articulate a way forward.”

—Lawrence Tabak,
National Institutes of Health

The NIH is calling on other stakeholders to help tackle the problem, including journal editors and reviewers, and university promotion and tenure committees. According to Collins and Tabak, the pressures for rapid publication and the ties of high-impact publications to academic career progression have contributed to the poor rates of reproducibility.

“It’s not NIH’s problem to fix this alone. The first burden is also on us academics,” said Daria Mochly-Rosen, who is a professor of translational medicine in the Department of Chemical Systems and Biology at the **Stanford University School of Medicine**. She also was cofounder and CEO of Kai Pharmaceuticals Inc., which was acquired by Amgen in 2012, and is cofounder of **Aldea Pharmaceuticals Inc.**

Although the proposed NIH training module is expected to launch at year end, Mochly-Rosen and others at Stanford have already started implementing procedures for better training of lab scientists who are conducting experiments.

Often, she said, lack of reproducibility arises when experiments have confounding factors that are not perceived as important.

For example, cell lines might have guidelines for use between specific passage numbers. Those guidelines can often inadvertently get lost as the cells are used by different laboratory researchers, leading to contradictory results between labs.

In other situations, one lab might perform animal experiments at one time of day and a second lab might work at a different time of day. If circadian rhythms influence the system under investigation, the data may not be replicated in the repeat experiment.

Mochly-Rosen said that reproducibility problems can be partly mitigated by better recording of experimental conditions. She does not require the fully detailed forms used in industry but does encourage her lab members to record information on experimental conditions, reagent source and lot numbers and any deviations from the core protocol.

To ensure blinding and maintain objectivity, results are analyzed independently by another lab member who is blinded to the experimental conditions.

“This costs a lot of goodwill—you’re asking someone to give a day of their time to do something that they get no credit for. But out of a sense of community people will do this,” she said.

Prior to submitting a publication, Mochly-Rosen requests that some experiments be repeated by an independent scientist based only on the information included in the manuscript to ensure that sufficient detail is provided to reproduce the data.

Mochly-Rosen said that journals and reviewers also have a responsibility to maintain scientific standards. “It’s unacceptable to allow data to be published without including both positive and negative controls,” she said.

Indeed, the NIH is calling on journals to help raise standards and wants widespread adoption of practices already followed by **Nature Publishing Group** and the journals of the **American Association for the**

Advancement of Science. These include an expanded online methods section of supplementary material and checklists for editors and reviewers to ensure critical experimental design features are included.

Collins and Tabak also recommend that journals devote more space to reporting negative findings and corrections to earlier work.

Ellis told *SciBX* that publishing negative findings is particularly important and could help researchers avoid wasting resources and time on pursuing fruitless avenues already explored by others.

He outlined three types of negative data that should be published more widely: studies on hypotheses that prove to be incorrect; studies that try but fail to reproduce work published by someone else; and some of the scientist’s own studies that did not work in the context of a larger study that was consistent with the principal investigator’s hypothesis.

For example, he said, if an experiment did not work in 10% serum but did work in 1% serum, that information should be included.

“Reviewers only want the perfect story, but we should be able to publish all the relevant information, both positive and negative,” he said.

Mochly-Rosen said that many researchers are introducing better practices for ensuring reproducibility in their labs. However, she said, talking directly with students and colleagues can be difficult. Researchers sometimes react negatively to the issue as they think it impugns their scientific ability or integrity.

“The topic has been taboo, but we need to address it. This is not about improving the robustness of our published work for industry. We should do it because it’s the right thing to do,” she said.

Fishburn, C.S. *SciBX* 7(10); doi:10.1038/scibx.2014.275
Published online March 13, 2014

REFERENCES

1. Collins, F.S. & Tabak, L.A. *Nature* **505**, 612–613 (2014)
2. Prinz, F. *et al. Nat. Rev. Drug Discov.* **10**, 712 (2011)
3. Begley, C.G. & Ellis, L.M. *Nature* **483**, 531–533 (2012)
4. Fulmer, T. *SciBX* **5**(34); doi:10.1038/scibx.2012.888
5. Haas, M.J. *SciBX* **7**(3); doi:10.1038/scibx.2014.76

COMPANIES AND INSTITUTIONS MENTIONED

Aldea Pharmaceuticals Inc., Redwood City, Calif.
American Association for the Advancement of Science, Washington, D.C.
Amgen Inc. (NASDAQ:AMGN), Thousand Oaks, Calif.
Bayer AG (Xetra:BAYN), Leverkusen, Germany
figshare, London, U.K.
Global Biological Standards Institute, Washington, D.C.
National Institutes of Health, Bethesda, Md.
Nature Publishing Group, London, U.K.
PLOS, San Francisco, Calif.
Science Exchange, Palo Alto, Calif.
Stanford University School of Medicine, Stanford, Calif.
SWOG, Portland, Ore.
The University of Texas MD Anderson Cancer Center, Houston, Texas