



NIH plans to enhance reproducibility

Francis S. Collins and Lawrence A. Tabak discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

A growing chorus of concern, from scientists and laypeople, contends that the complex system for ensuring the reproducibility of biomedical research is failing and is in need of restructuring^{1,2}. As leaders of the US National Institutes of Health (NIH), we share this concern and here explore some of the significant interventions that we are planning.

Science has long been regarded as 'self-correcting', given that it is founded on the replication of earlier work. Over the long term, that principle remains true. In the

shorter term, however, the checks and balances that once ensured scientific fidelity have been hobbled. This has compromised the ability of today's researchers to reproduce others' findings.

Let's be clear: with rare exceptions, we have no evidence to suggest that irreproducibility is caused by scientific misconduct. In 2011, the Office of Research Integrity of the US Department of Health and Human Services pursued only 12 such cases³. Even if this represents only a fraction of the actual problem, fraudulent papers are vastly

outnumbered by the hundreds of thousands published each year in good faith.

Instead, a complex array of other factors seems to have contributed to the lack of reproducibility. Factors include poor training of researchers in experimental design; increased emphasis on making provocative statements rather than presenting technical details; and publications that do not report basic elements of experimental design⁴. Crucial experimental design elements that are all too frequently ignored include blinding, randomization, replication, sample-size calculation and the effect of sex differences.

And some scientists reputedly use a 'secret sauce' to make their experiments work — and withhold details from publication or describe them only vaguely to retain a competitive edge⁵. What hope is there that other scientists will be able to build on such work to further biomedical progress?

Exacerbating this situation are the policies and attitudes of funding agencies, academic centres and scientific publishers. Funding agencies often uncritically encourage the overvaluation of research published in high-profile journals. Some academic centres also provide incentives for publications in such journals, including promotion and tenure, and in extreme circumstances, cash rewards⁶.

Then there is the problem of what is not published. There are few venues for researchers to publish negative data or papers that point out scientific flaws in previously published work. Further compounding the problem is the difficulty of accessing unpublished data — and the failure of funding agencies to establish or enforce policies that insist on data access.

PRECLINICAL PROBLEMS

Reproducibility is potentially a problem in all scientific disciplines. However, human clinical trials seem to be less at risk because they are already governed by various regulations that stipulate rigorous design and independent oversight — including randomization, blinding, power estimates, pre-registration of outcome measures in standardized, public databases such as ClinicalTrials.gov and oversight by institutional review boards and data safety monitoring boards. Furthermore, the clinical trials community has taken important steps towards adopting standard reporting elements⁷.

Preclinical research, especially work that uses animal models¹, seems to be the area that is currently most susceptible to reproducibility issues. Many of these failures have simple and practical explanations: different animal strains, different lab environments or subtle changes in protocol. Some irreproducible reports are probably the result of coincidental findings that happen to reach statistical significance, coupled with publication bias.

Another pitfall is overinterpretation of creative 'hypothesis-generating' experiments, which are designed to uncover new avenues of inquiry rather than to provide definitive proof for any single question. Still, there remains a troubling frequency of published reports that claim a significant result, but fail to be reproducible.

PROPOSED NIH ACTIONS

As a funding agency, the NIH is deeply concerned about this problem. Because poor training is probably responsible for at least some of the challenges, the NIH is developing a training module on enhancing reproducibility and transparency of research findings, with an emphasis on good experimental design. This will be incorporated into the mandatory training on responsible conduct of research for NIH intramural postdoctoral fellows later this year. Informed by this pilot, final materials will be posted on the NIH website by the end of this year for broad dissemination, adoption or adaptation, on the basis of local institutional needs.

Several of the NIH's institutes and centres are also testing the use of a checklist to ensure a more systematic evaluation of grant applications. Reviewers are reminded to check, for example, that appropriate experimental design features have been addressed, such as an analytical plan, plans for randomization, blinding and so on. A pilot was launched last year that we plan to complete by the end of this year to assess the value of assigning at least one reviewer on each panel the specific task of evaluating the 'scientific premise' of the application: the key publications on which the application is based (which may or may not come from the applicant's own research efforts). This question will be particularly important when a potentially costly human clinical trial is proposed, based on animal-model results. If the antecedent work is questionable and the trial is particularly important, key preclinical studies may first need to be validated independently.

Informed by feedback from these pilots, the NIH leadership will decide by the fourth quarter of this year which approaches to adopt agency-wide, which should remain specific to institutes and centres, and which to abandon.

The NIH is also exploring ways to provide greater transparency of the data that are the basis of published manuscripts. As part of our Big Data initiative, the NIH has requested applications to develop a Data Discovery Index (DDI) to allow investigators to locate and access unpublished, primary data (see go.nature.com/rjjfoj). Should an investigator use these data in new work, the owner of the data set could be cited, thereby creating a new metric of scientific contribution unrelated

to journal publication, such as downloads of the primary data set. If sufficiently meritorious applications to develop the DDI are received, a funding award of up to three years in duration will be made by September 2014. Finally, in mid-December, the NIH launched an online forum called PubMed Commons (see go.nature.com/8m4pfp) for open discourse about published articles. Authors can join and rate or contribute comments, and the system is being evaluated and refined in the coming months. More than 2,000 authors have joined to date, contributing more than 700 comments.

COMMUNITY RESPONSIBILITY

Clearly, reproducibility is not a problem that the NIH can tackle alone. Consequently, we are reaching out broadly to the research community, scientific publishers, universities, industry, professional organizations, patient-advocacy groups and other stakeholders to take the steps necessary to reset the self-corrective process of scientific inquiry. Journals should be encouraged to devote more space to research conducted in an exemplary manner that reports negative findings, and should make room for papers that correct earlier work.

We are pleased to see that some of the leading journals have begun to change their review practices. For example, Nature Publishing Group, the publishers of this journal, announced⁸ in May 2013 the following: restrictions on the length of methods sections have been abolished to ensure the reporting of key methodological details; authors use a checklist to facilitate the verification by editors and reviewers that critical experimental design features have been incorporated into the report, and editors scrutinize the statistical treatment of the studies reported more thoroughly with the help of statisticians. Furthermore, authors are encouraged to provide more raw data to accompany their papers online.

Similar requirements have been implemented by the journals of the American Association for the Advancement of Science — *Science Translational Medicine* in 2013 and *Science* earlier this month⁹ — on the basis of, in part, the efforts of the NIH's National Institute of Neurological Disorders and Stroke to increase the transparency of how work is conducted¹⁰.

Perhaps the most vexed issue is the academic incentive system. It currently over-emphasizes publishing in high-profile journals. No doubt worsened by current budgetary woes, this encourages rapid submission of research findings to the detriment of careful replication. To address this, the NIH is contemplating modifying the format of its 'biographical sketch' form, which grant applicants are required to complete, to emphasize the significance

of advances resulting from work in which the applicant participated, and to delineate the part played by the applicant. Other organizations such as the Howard Hughes Medical Institute have used this format and found it more revealing of actual contributions to science than the traditional list of unannotated publications. The NIH is also considering providing greater stability for investigators at certain, discrete career stages, utilizing grant mechanisms that

"Efforts by the NIH alone will not be sufficient to effect real change in this unhealthy environment."

allow more flexibility and a longer period than the current average of approximately four years of support per project.

In addition, the NIH is examining ways to anonymize the peer-review process

to reduce the effect of unconscious bias (see go.nature.com/g5xr3c). Currently, the identifiers and accomplishments of all research participants are known to the reviewers. The committee will report its recommendations within 18 months.

Efforts by the NIH alone will not be sufficient to effect real change in this unhealthy environment. University promotion and tenure committees must resist the temptation to use arbitrary surrogates, such as the number of publications in journals with high impact factors, when evaluating an investigator's scientific contributions and future potential.

The recent evidence showing the irreproducibility of significant numbers of biomedical-research publications demands immediate and substantive action. The NIH is firmly committed to making systematic changes that should reduce the frequency and severity of this problem — but success will come only with the full engagement of the entire biomedical-research enterprise. ■

Francis S. Collins is director and **Lawrence A. Tabak** is principal deputy director of the US National Institutes of Health, Bethesda, Maryland, USA. e-mail: lawrence.tabak@nih.gov

1. Prinz, F., Schlange, T. & Asadullah, K. *Nature Rev. Drug Disc.* **10**, 712–713 (2011).
2. *The Economist* 'Trouble at the Lab' (19 October 2013); available at <http://go.nature.com/dstij3>.
3. US Department of Health and Human Services, 2011 *Office of Research Integrity Annual Report 2011* (US HHS, 2011); available at <http://go.nature.com/t7ykv>.
4. Carp, J. *NeuroImage* **63**, 289–300 (2012).
5. Vasilevsky, N. A. et al. *PeerJ* **1**, e148 (2013).
6. Franzoni, C., Scellato, G. & Stephan, P. *Science* **333**, 702–703 (2011).
7. Moher, D., Jones, A. & Lepage, L. for the CONSORT Group *J. Am. Med. Assoc.* **285**, 1992–1995 (2001).
8. *Nature* **496**, 398 (2013).
9. McNutt, M. *Science* **343**, 229 (2014).
10. Landis, S. C. et al. *Nature* **490**, 187–191 (2012).