New NIH 2-page Rigor and Transparency Document

Grants affected in 2016 include:

NOT-OD-16-081
NOT-OD-16-058
NOT-OD-16-034
NOT-OD-16-011
NOT-OD-16-031
NOT-OD-16-012
NOT-OD-16-005
NOT-OD-16-004
NOT-OD-15-103
NOT-OD-15-102

Objectives of this presentation

- Describe reproducibility challenges and NIH focus
- Summarize changes to application instructions and review criteria
- Useful resources to help researchers in meeting criteria

Research reproducibility is a challenge: Funding agencies and popular press weigh in

PERSPECTIVE

doi:10.1038/nature11556

A call for transparent reporting to optimize the predictive value of preclinical research

Story C. Landis[1], Susan G. Amara[2], Khusru Asadullah[3], Chris P. Austin[4], Robi Blumenstein[5], Eileen W. Bradley[6], Ronald G. Crystal[7], Robert B. Darnell[8], Robert J. Ferrante[9], Howard Fillit[10], Robert Finkelstein[1], Marc Fisher[11], Howard E. Gendelman[12], Robert M. Golub[13], John L. Goudreau[14], Robert A. Gross[15], Amelie K. Gubitz[1], Sharon E. Hesterlee[16], David W. Howells[17], John Huguenard[18], Katrina Kelner[19], Walter Koroshetz[1], Dimitri Krainc[20], Stanley E. Lazic[21], Michael S. Levine[22], Malcolm R. Macleod[23], John M. McCall[24], Richard T. Moxley III[25], Kalyani Narasimhan[26], Linda J. Noble[27], Steve Perrin[28], John D. Porter[1], Oswald Steward[29], Ellis Unger[30], Ursula Utz[1] & Shai D. Silberberg[1]

The US National Institute of Neurological Disorders and Stroke convened major stakeholders in June 2012 to discuss how to improve the methodological reporting of animal studies in grant applications and publications. The main workshop recommendation is that at a minimum studies should report on sample-size estimation, whether and how animals were randomized, whether investigators were blind to the treatment, and the handling of data. We recognize that achieving a meaningful improvement in the quality of reporting will require a concerted effort by investigators, reviewers, funding agencies and journal editors. Requiring better reporting of animal studies will raise awareness of the importance of rigorous study design to accelerate scientific progress.

Good experimental design is nothing new

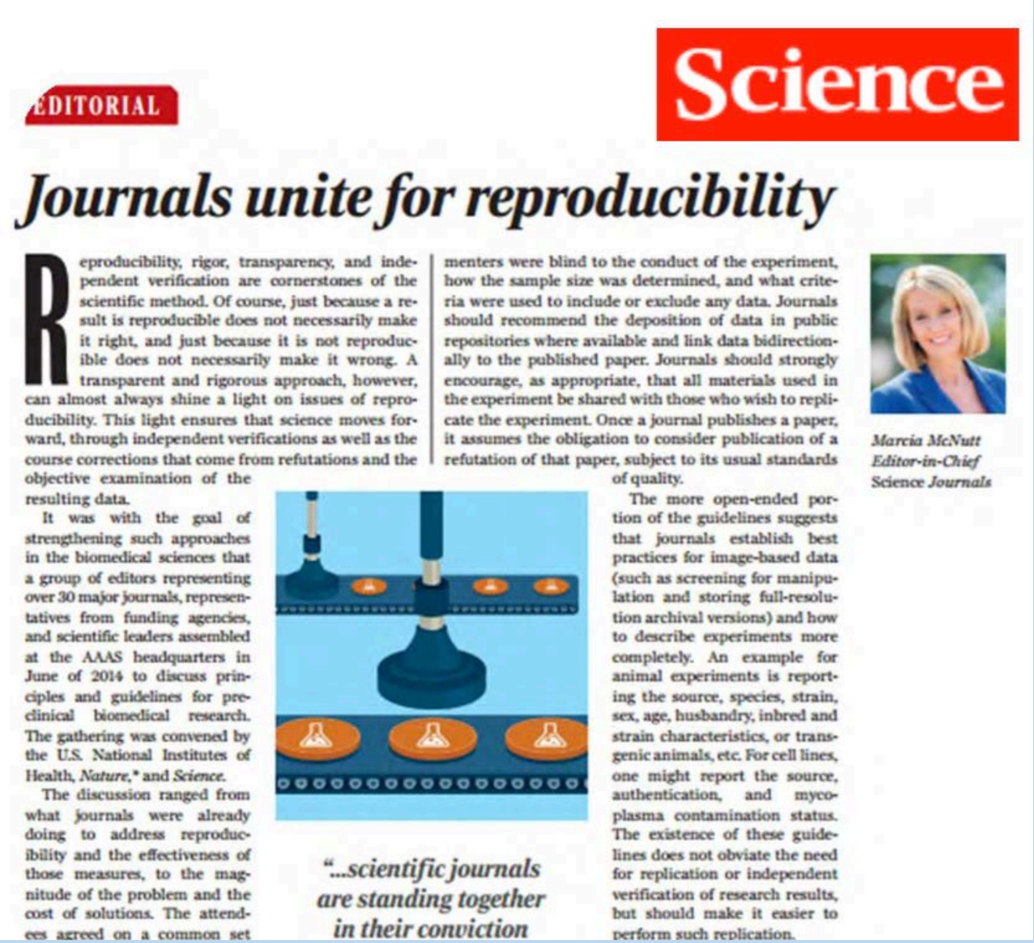
"There's a flaw in your experimental design. All the mice are scorpios."

Five requirements for a "good" experimental design:

- Be unbiased
- Have high precision
- Have a wide range of applicability
- Be simple
- Have the ability to calculate uncertainty

COX, D.R. *Planning Experiments*, John Wiley and Sons, New York, 1958.

Returning to good design principles to enhance reproducibility is important to NIH and top journals



EDITORIAL

Science

Journals unite for reproducibility

Reproducibility, rigor, transparency, and independent verification are cornerstones of the scientific method. Of course, just because a result is reproducible does not necessarily make it right, and just because it is not reproducible does not necessarily make it wrong. A transparent and rigorous approach, however, can almost always shine a light on issues of reproducibility. This light ensures that science moves forward, through independent verifications as well as the course corrections that come from refutations and the objective examination of the resulting data.

It was with the goal of strengthening such approaches in the biomedical sciences that a group of editors representing over 30 major journals, representatives from funding agencies, and scientific leaders assembled at the AAAS headquarters in June of 2014 to discuss principles and guidelines for preclinical biomedical research. The gathering was convened by the U.S. National Institutes of Health, *Nature*,* and *Science*.

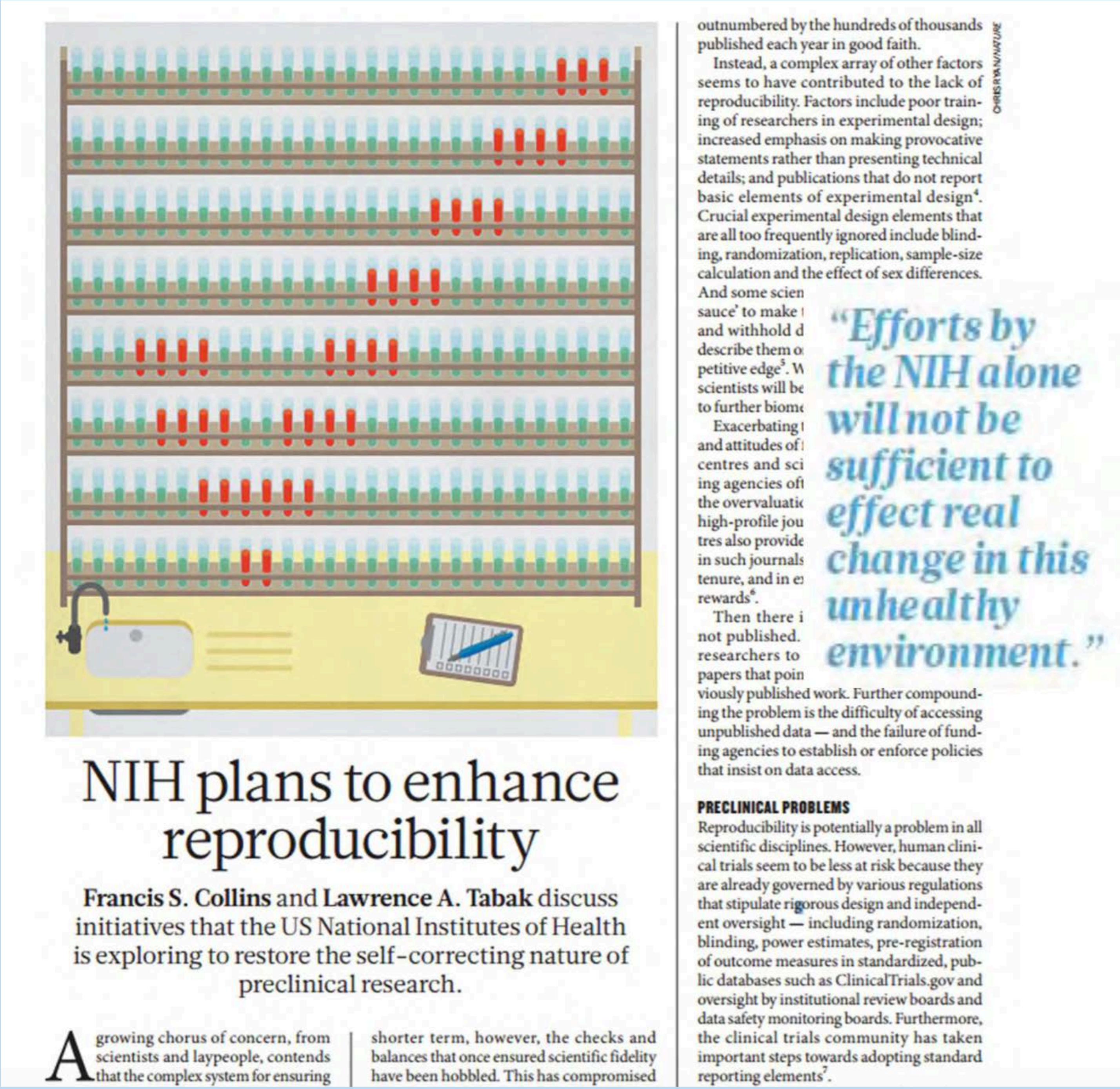
The discussion ranged from what journals were already doing to address reproducibility and the effectiveness of those measures, to the magnitude of the problem and the cost of solutions. The attendees agreed on a common set

menters were blind to the conduct of the experiment, how the sample size was determined, and what criteria were used to include or exclude any data. Journals should recommend the deposition of data in public repositories where available and link data bidirectionally to the published paper. Journals should strongly encourage, as appropriate, that all materials used in the experiment be shared with those who wish to replicate the experiment. Once a journal publishes a paper, it assumes the obligation to consider publication of a refutation of that paper, subject to its usual standards of quality.

The more open-ended portion of the guidelines suggests that journals establish best practices for image-based data (such as screening for manipulation and storing full-resolution archival versions) and how to describe experiments more completely. An example for animal experiments is reporting the source, species, strain, sex, age, husbandry, inbred and strain characteristics, or transgenic animals, etc. For cell lines, one might report the source, authentication, and mycoplasma contamination status. The existence of these guidelines does not obviate the need for replication or independent verification of research results, but should make it easier to perform such replication.

"...scientific journals are standing together in their conviction

Marcia McNutt
Editor-in-Chief
Science Journals



CHRIS RYAN/NATURE

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

shorter term, however, the checks and balances that once ensured scientific fidelity have been hobbled. This has compromised

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 design[4]. 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 scien sauce' to make and withhold d describe them o petitive edge[5]. W scientists will be to further biome

Exacerbating and attitudes of centres and sci ing agencies oft the overvaluati high-profile jou tres also provide in such journals tenure, and in e rewards[6].

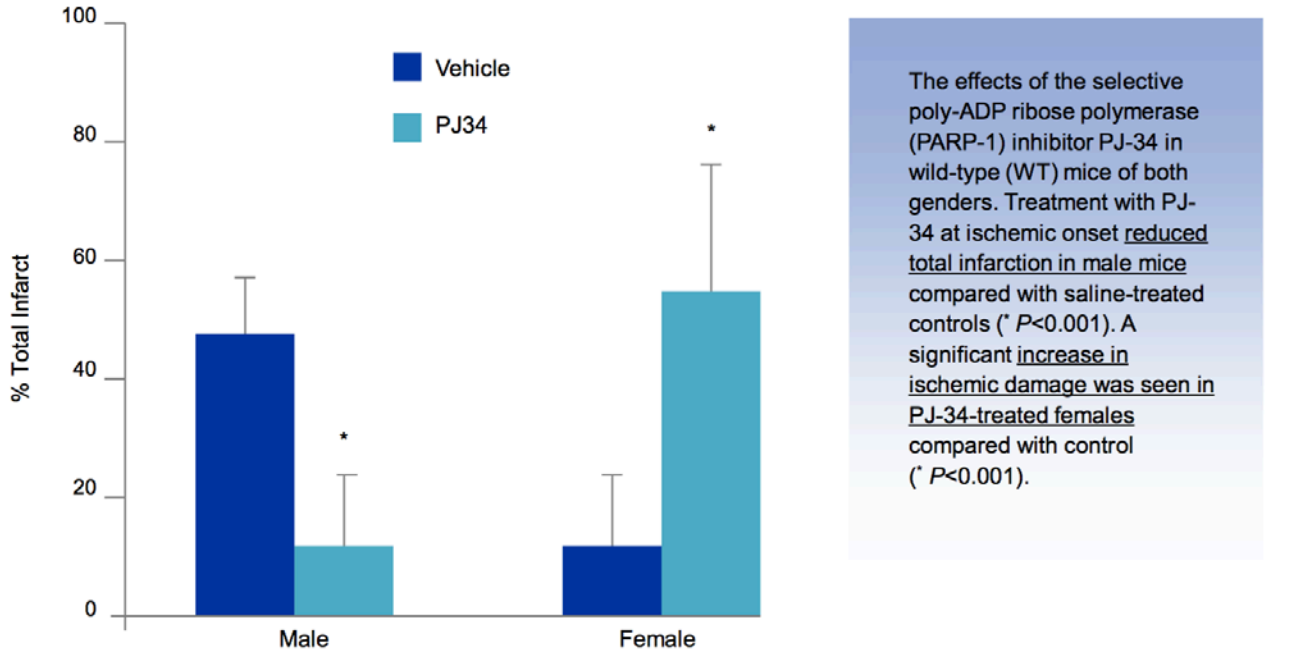
Then there i not published. researchers to papers that poin viously published work. Further compounding the problem is the difficulty of accessing unpublished data — and the failure of funding agencies to establish or enforce policies that insist on data access.

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

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 elements[7].

Specifics called out by NIH: **Sex** of subjects



The effects of the selective poly-ADP ribose polymerase (PARP-1) inhibitor PJ-34 in wild-type (WT) mice of both genders. Treatment with PJ-34 at ischemic onset reduced total infarction in male mice compared with saline-treated controls (* $P<0.001$). A significant increase in ischemic damage was seen in PJ-34-treated females compared with control (* $P<0.001$).

McCullough LD, et al. *Journal of Cerebral Blood Flow & Metabolism* (2005) **25**, 502 512.

NIH National Institutes of Health Office of Extramural Research

COMMENT

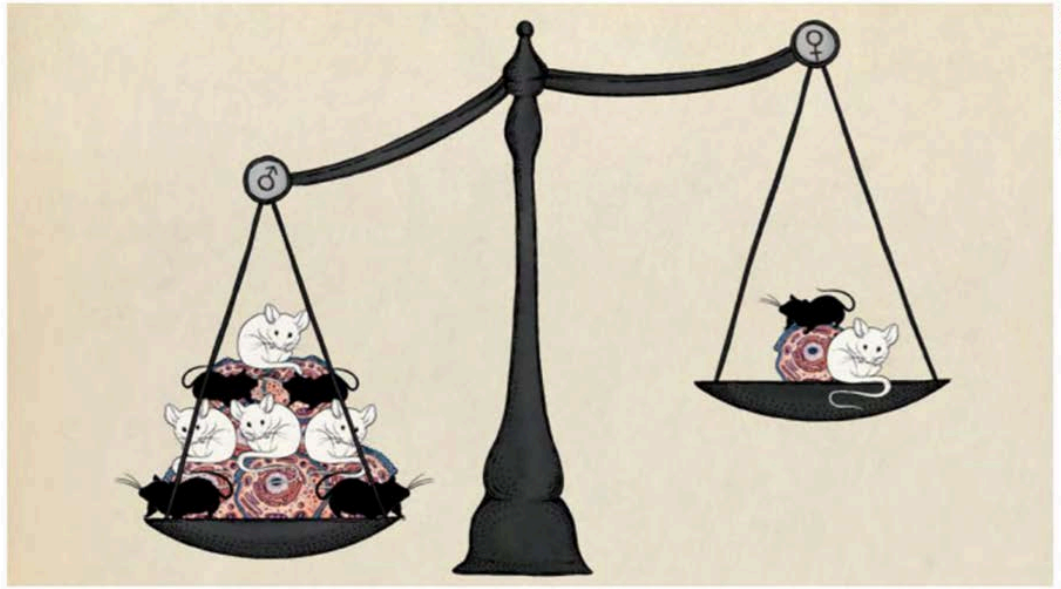


ILLUSTRATION BY KATE SCOTT

NIH to balance sex in cell and animal studies

Janine A. Clayton and **Francis S. Collins** unveil policies to ensure that preclinical research funded by the US National Institutes of Health considers females and males.

More than two decades ago, the US National Institutes of Health (NIH) established the Office of Research on Women's Health (ORWH). At that time, the Congressional Caucus for Women's Issues, women's health advocacy groups and NIH scientists and leaders agreed that excluding women from clinical research was bad for women and bad for science. In 1993, the NIH Revitalization Act required the inclusion of women in NIH-funded clinical research.

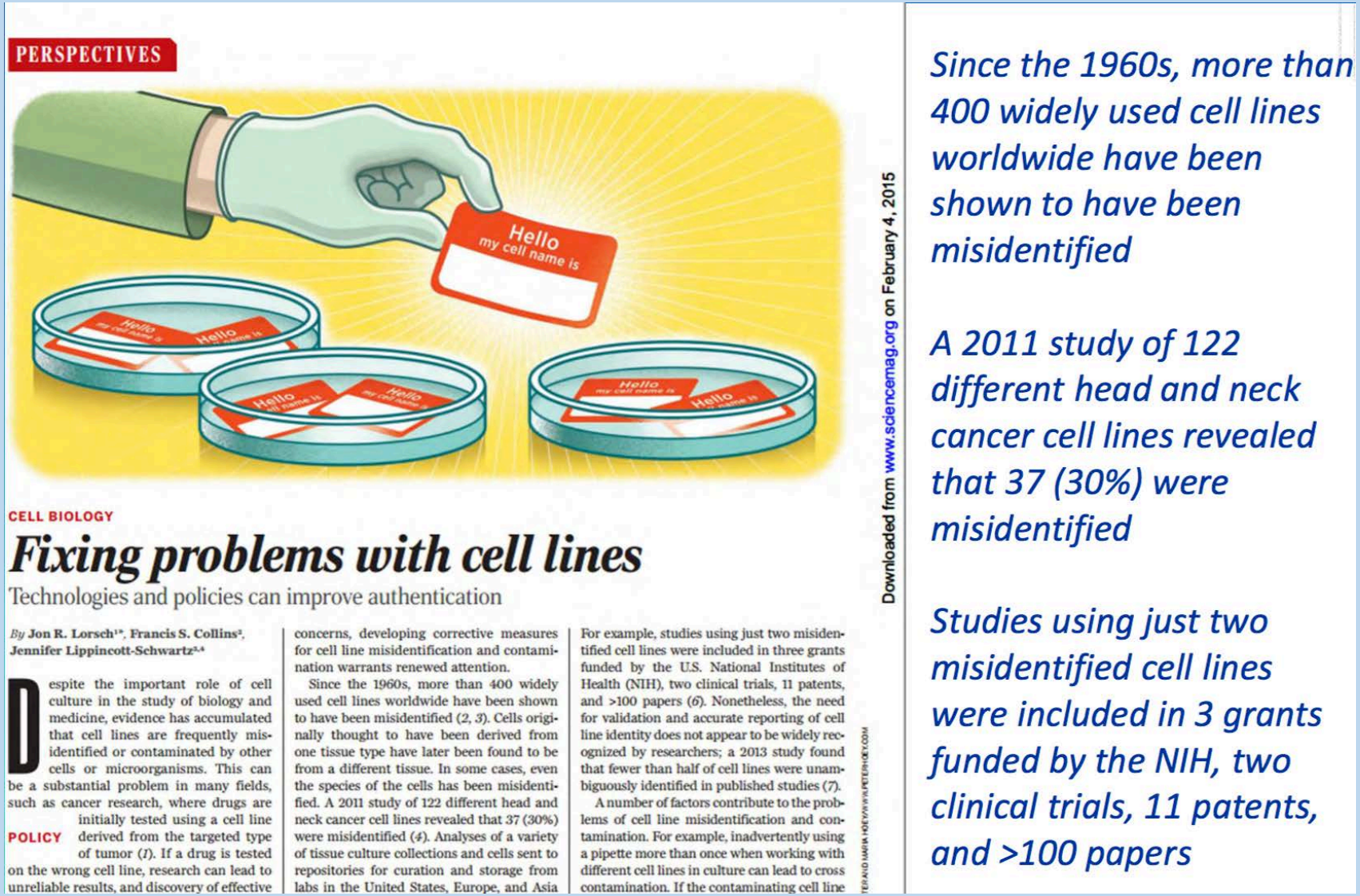
Today, just over half of NIH-funded clinical-research participants are women. We know much more about the role of sex and gender in medicine, such as that low-dose aspirin has different preventive effects in women and men, and that drugs such as calls to action[1]. Publications often continue to neglect sex-based considerations and analyses in preclinical studies[2,3]. Reviewers, for the most part, are not attuned to this failure. The over-reliance on male animals and cells in preclinical research obscures key sex differences that could guide clinical studies. And it might be harmful: women experience higher rates of adverse drug reactions than men do[4]. Furthermore, inadequate inclusion of female cells and animals in experiments and inadequate analysis of data by sex may well contribute to the troubling rise of irreproducibility in preclinical biomedical research, which the NIH is now actively working to address[5,6].

The NIH plans to address the issue of sex and gender inclusion across biomedical research multi-dimen- stakeholders including publishers. This move is essential, potentially very powerful and need not be difficult or costly.

BETTER WITH BOTH

Certain rigorous studies evaluating the effects of sex differences have been effective in bridging the divide between animal and human work. One example concerns multiple sclerosis (MS). Women are more susceptible to MS than men are, but develop less-severe forms of the disease. The most widely accepted MS animal model — rodent experimental autoimmune encephalomyelitis (EAE) — has revealed[7] that sex differences in MS are related to both reproductive and non-reproductive factors. Findings[8] that oestrogen therapy provided benefits in rodent EAE

Specifics called out by NIH: **Cell line** identity



PERSPECTIVES

CELL BIOLOGY

Fixing problems with cell lines

Technologies and policies can improve authentication

By **Jon R. Lorsch[1*], Francis S. Collins[2], Jennifer Lippincott-Schwartz[3,4]**

Despite the important role of cell culture in the study of biology and medicine, evidence has accumulated that cell lines are frequently misidentified or contaminated by other cells or microorganisms. This can be a substantial problem in many fields, such as cancer research, where drugs are initially tested using a cell line derived from the targeted type of tumor (*1*). If a drug is tested on the wrong cell line, research can lead to unreliable results, and discovery of effective

POLICY

concerns, developing corrective measures for cell line misidentification and contamination warrants renewed attention.

Since the 1960s, more than 400 widely used cell lines worldwide have been shown to have been misidentified (*2*, *3*). Cells originally thought to have been derived from one tissue type have later been found to be from a different tissue. In some cases, even the species of the cells has been misidentified. A 2011 study of 122 different head and neck cancer cell lines revealed that 37 (30%) were misidentified (*4*). Analyses of a variety of tissue culture collections and cells sent to repositories for curation and storage from labs in the United States, Europe, and Asia

For example, studies using just two misidentified cell lines were included in three grants funded by the U.S. National Institutes of Health (NIH), two clinical trials, 11 patents, and >100 papers (*6*). Nonetheless, the need for validation and accurate reporting of cell line identity does not appear to be widely recognized by researchers; a 2013 study found that fewer than half of cell lines were unambiguously identified in published studies (*7*).

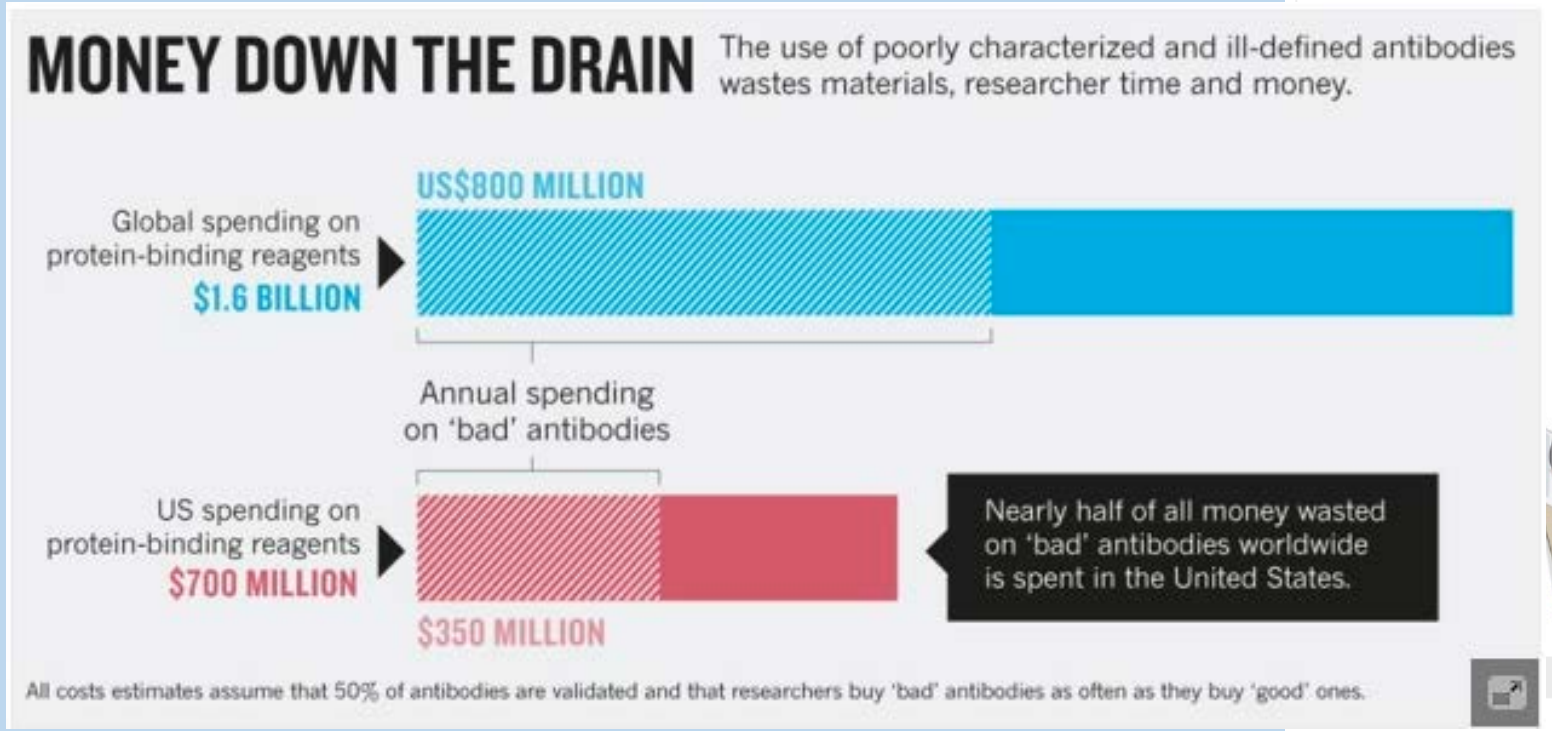
A number of factors contribute to the problems of cell line misidentification and contamination. For example, inadvertently using a pipette more than once when working with different cell lines in culture can lead to cross contamination. If the contaminating cell line

Since the 1960s, more than 400 widely used cell lines worldwide have been shown to have been misidentified

A 2011 study of 122 different head and neck cancer cell lines revealed that 37 (30%) were misidentified

Studies using just two misidentified cell lines were included in 3 grants funded by the NIH, two clinical trials, 11 patents, and >100 papers

Specifics called out by NIH: **Antibodies**



NATURE | NEWS FEATURE

Reproducibility crisis: Blame it on the antibodies

Antibodies are the workhorses of biological experiments, but they are littering the field with false findings. A few evangelists are pushing for change.

Monya Baker

19 May 2015

PDF | Rights & Permissions

Illustration by Nik Spencer/Nature

In 2006, things were looking pretty good for David Rimm, a pathologist at Yale University in New Haven, Connecticut. He had developed a test to guide effective treatment of the skin cancer melanoma, and it promised to save lives. It relied on antibodies — large, Y-shaped proteins that

Specifics called out by NIH: **Statistics**

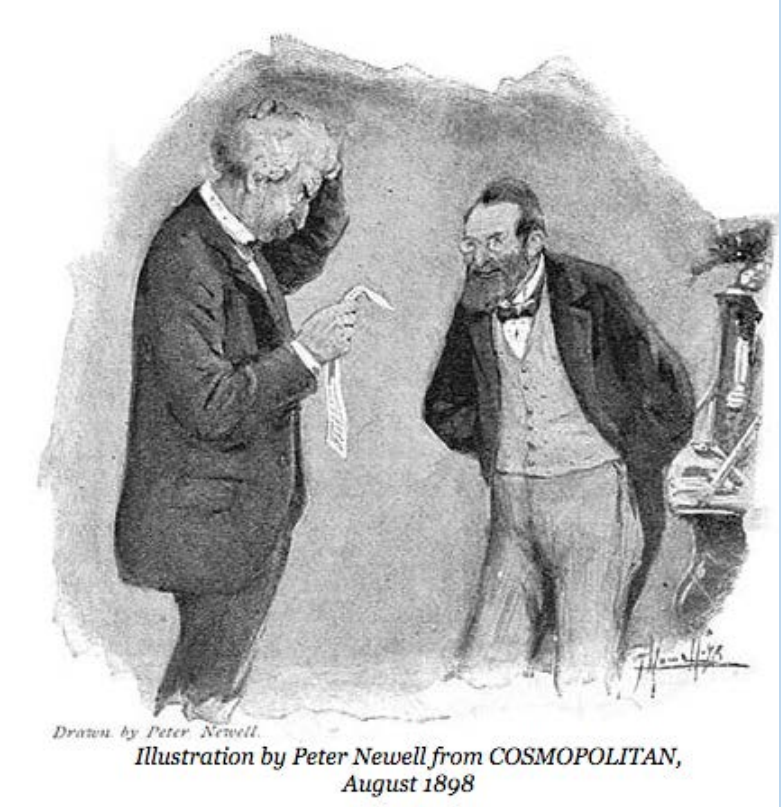


Illustration by Peter Newell from COSMOPOLITAN, August 1898

"There are three kinds of lies: lies, damned lies and statistics."
- *possibly Mark Twain*

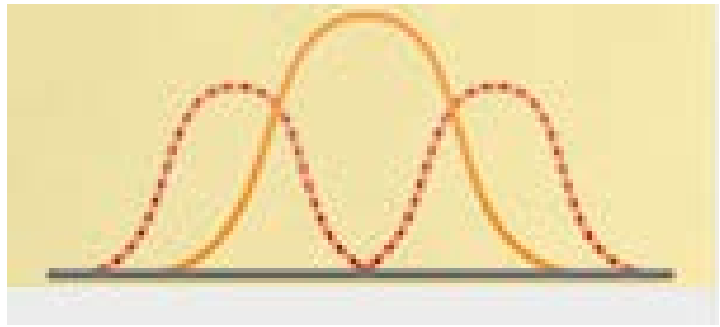
NATURE | NEWS

Statisticians issue warning over misuse of *P* values

Policy statement aims to halt missteps in the quest for certainty.

Monya Baker

07 March 2016



PDF | Rights & Permissions

Misuse of the *P* value — a common test for judging the strength of scientific evidence — is contributing to the number of research findings that cannot be reproduced, the American Statistical Association (ASA) warns in a statement released today[1]. The group has taken the unusual step of issuing principles to guide use of the *P* value, which it says cannot determine whether a hypothesis is true or whether results are important.

This is the first time that the 177-year-old ASA has made explicit recommendations on such a foundational matter in statistics, says executive director Ron Wasserstein. The society's members had become increasingly concerned that the *P* value was being misapplied in ways that cast doubt on statistics generally, he adds.

NIH focus areas in the coming years will be a series of pilots, that change NIH grant criteria/types

Pilot Focus	Types of Efforts Being Developed
Evaluation of scientific premise in grant applications	New Funding Opportunities with additional review criteria regarding scientific premise
Checklist and Reporting Guidelines	Reviewer checklists regarding reporting standards and scientific rigor
Changes to Biosketch	Biosketch pilot with focus on accomplishments and not just publications
Approaches to reduce "perverse incentives to publish	Exploring award options with a longer period of support for investigators
Supporting replication studies	New Funding Opportunities for replication studies, and options to assess whether pre-clinical findings should be replicated
Training	Developing materials on experimental design
Other efforts	Use of Prize Challenges to encourage reproducibility of results, PubMed Commons

Focus of this Presentation

Principles and guidelines for reporting preclinical research

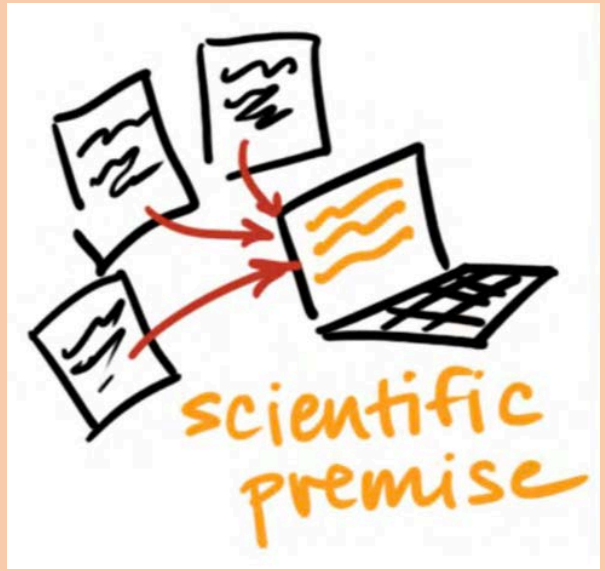
- Rigorous statistical analysis
- Transparency in reporting
  - Standards
  - Replicates
  - Statistics
  - Randomization
  - Blinding
  - Sample size estimation
  - Inclusion/exclusion criteria
- Data and material sharing
- Consideration of refutations
- Consider establishing best practices for:
  - Antibodies
  - Cell lines
  - Animals

http://www.nih.gov/about/reporting-preclinical-research.htm

Review criteria changes: What will be scored and how will it be scored?

Element of Rigor	Section of Application	Criterion Score	Additional Review Consideration	Contribute to Overall Impact?
Scientific Premise		Significance	NA	Yes
Scientific Rigor	Research Strategy	Approach	NA	Yes
Consideration of Sex and Other Relevant Biological Variables		Approach	NA	Yes
Authentication of Key Biological and/or Chemical Resources	New Attachment	NA	Acceptable or unacceptable	No

Review criteria changes:
Scientific **Premise**: contributes directly to overall score



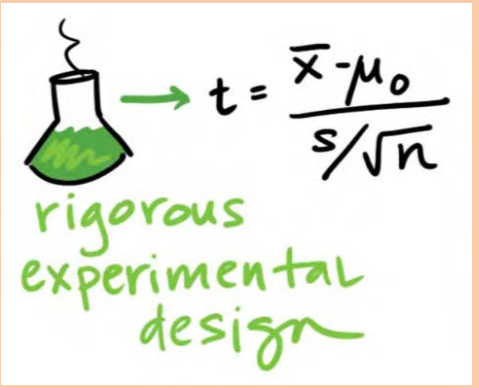
- **All research builds on prior research**, whether observations, preliminary data or published literature. The scientific premise for an application is the research that is used to form the basis for the proposed research.
- **NIH expects applicants to describe the general strengths and weaknesses of the prior research** being cited by the applicant as crucial to support the application. It is expected that this consideration of general strengths and weaknesses could include attention to the rigor of the previous experimental designs, as well as the incorporation of relevant biological variables and **authentication of key resources**.

Review criteria changes: Scientific **Premise**: review questions

- Does the project address an important problem or critical barrier to progress in the field?
- Is there a strong scientific premise for the project?
- If the aims of the project are achieved, how will scientific knowledge, technical capability, and/or clinical practice be improved?
- How will successful completion of the aims change the concepts, methods, technologies, treatments, services, or preventative interventions that drive this field?

http://grants.nih.gov/reproducibility/faqs.htm#4825

Review criteria changes: Scientific **Rigor**: contributes to overall score



- **Scientific rigor** is the strict application of the scientific method to ensure robust and unbiased experimental design, methodology, analysis, interpretation and reporting of results.
- NIH expects applicants to describe the experimental design and methods proposed and how they will achieve robust and unbiased results. Robust and unbiased results are obtained using methods designed to avoid bias and these results can be reproduced under well-controlled and reported experimental conditions.

Review criteria changes: Scientific **Rigor** approach: review questions

- Are the overall strategy, methodology, and analysis well-reasoned and appropriate to accomplish the specific aims of the project??
- *Have the investigators presented strategies to ensure a robust and unbiased approach, as appropriate for the work proposed?*
- Are potential problems, alternative strategies, and benchmarks for success presented?
- If the project is in the early stages of development, will the strategy establish feasibility and will particularly risky aspects be managed?

http://grants.nih.gov/reproducibility/faqs.htm#4825

Review criteria changes: Consideration of **Variables, e.g. Sex**: scored

- Biological variables, such as **sex**, **age**, **weight**, and underlying health conditions, are often critical factors affecting health or disease.
- NIH expects that sex as a biological variable will be factored into research designs, analyses, and reporting in vertebrate animal and human studies. Strong justification from the scientific literature, preliminary data or other relevant considerations must be provided for applications proposing to study only one sex.

Review criteria changes: Scientific **Variables**: review question

- *Have the investigators presented adequate plans to address relevant biological variable, such as sex, for studies in vertebrate animals or human subjects?*

http://grants.nih.gov/reproducibility/faqs.htm#4825

Review criteria changes: **Key Biological Resource** Authentication: pass/fail not directly added to score

- Key biological resources are: **antibodies, cell lines**, and other biologics that may vary from lab to lab and are the major source of error.
- What are the methods used to authenticate key resources, referencing relevant standards where applicable.
- If key resources have been purchased or obtained from an outside source that provided data on prior authentication, the investigator is still expected to provide their own authentication plans for these key resources.
- Actual data demonstrating that authenticated resources are available for the proposed research do not need to be included in the plan.

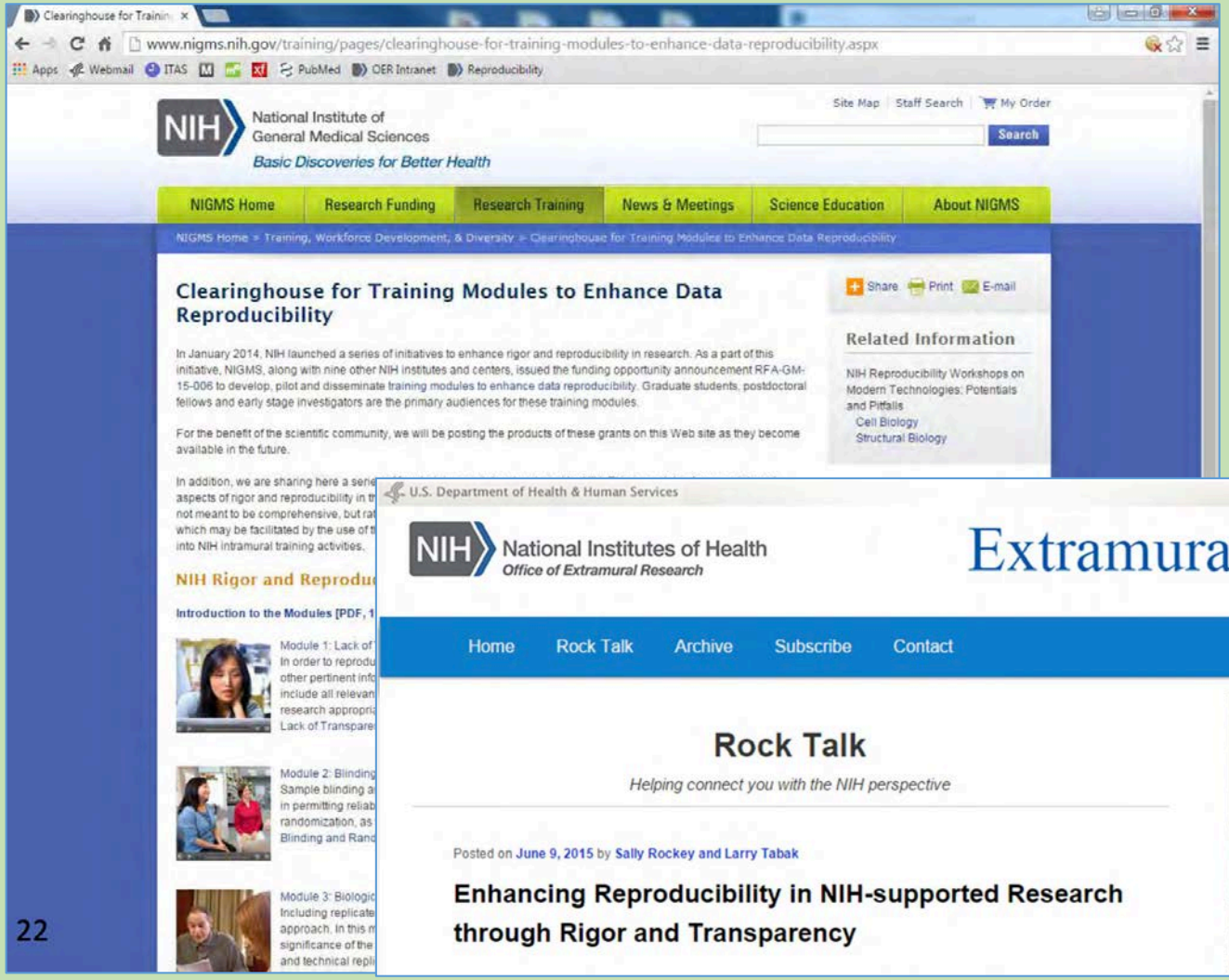
http://grants.nih.gov/reproducibility/faqs.htm#4847

Review criteria changes: **Key Biological Resource** Authentication: review question

- *For projects involving key biological and/or chemical resources, reviewers will comment on the brief plans proposed for identifying and ensuring the validity of those resources?*

http://grants.nih.gov/reproducibility/faqs.htm#4825

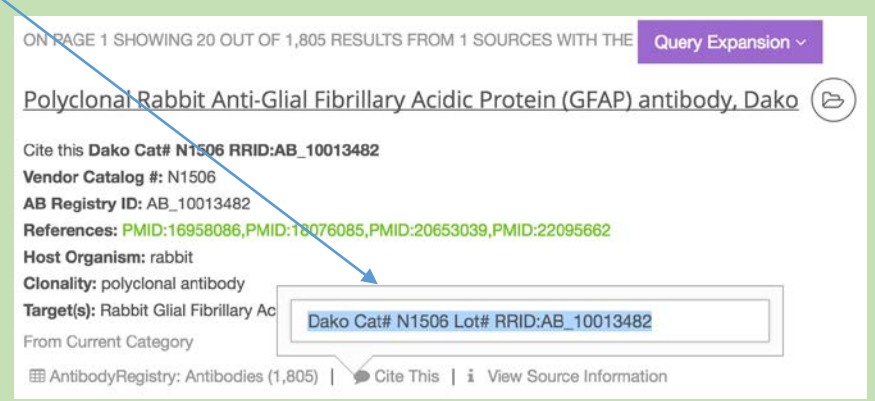
Resources for Investigators: NIH has many web pages devoted to understanding new criteria - we list the ones we felt most useful



Learning modules (videos) from NIH
https://loop.nigms.nih.gov/2015/03/clearinghouse-for-training-modules-to-enhance-data-reproducibility/
FAQs about reproducibility
http://grants.nih.gov/reproducibility/faqs.htm#V
Rigor and Reproducibility page
http://grants.nih.gov/reproducibility/index.htm

Authentication of Key Biological Resources: Places to look

- Identification of Cell Lines, Antibodies, Animals (from stock centers)
  - How to cite a particular resource: http://scicrunch.org/resources
    - Search for your resource and copy the information in the "cite this" button.



- Authenticate Cell Lines:
  - Lists known issues with a cell line, search Cellosaurus http://web.expasy.org/cellosaurus/
  - Checklist for validation of my cell line: ICLAC http://iclac.org/resources/cell-line-checklist/
    - Note, cell line authentication should be done at minimum at the beginning and end of the experiment and reported as part of the methods.
- Authenticate Antibodies:
  - Lists known issues with antibodies: AntibodyRegistry http://antibodyregistry.org/
  - Validation criteria for antibodies: Saper *et al.,* 2005 editorial http://www.ncbi.nlm.nih.gov/pubmed/16304632

Nature Reproducibility Special contains a nice synopsis of the key issues from statistics to antibodies

http://www.nature.com/news/reproducibility-1.17552

FEATURES

How quality control could save your science
It may not be sexy, but quality assurance is becoming a crucial part of lab life.
Nature (27 January 2016)

The biggest mystery in mathematics: Shinichi Mochizuki and the impenetrable proof
A Japanese mathematician claims to have solved one of the most important problems in his field. The trouble is, hardly anyone can work out whether he's right.
Nature (07 October 2015)

How scientists fool themselves – and how they can stop
Humans are remarkably good at self-deception. But growing concern about reproducibility is driving many researchers to seek ways to fight their own worst instincts.
Nature (07 October 2015)

Reproducibility crisis: Blame it on the antibodies
Antibodies are the workhorses of biological experiments, but they are littering the field with false findings. A few evangelists are pushing for change.
Nature (19 May 2015)

Statistical errors
P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume.
Nature (12 February 2014)

Replication studies: Bad copy
In the wake of high-profile controversies, psychologists are facing up to problems with replication.
Nature (16 May 2012)

NEWS AND ANALYSIS

Make journals report clinical trials properly
There is no excuse for the shoddy practice of allowing researchers to change outcomes and goals without saying so, says Ben Goldacre.
Nature (02 February 2016)