

# **NIH Perspectives on Enhancing Scientific Rigor and Reproducibility**

Michael S Lauer, MD

Deputy Director for Extramural Research, National Institutes of Health

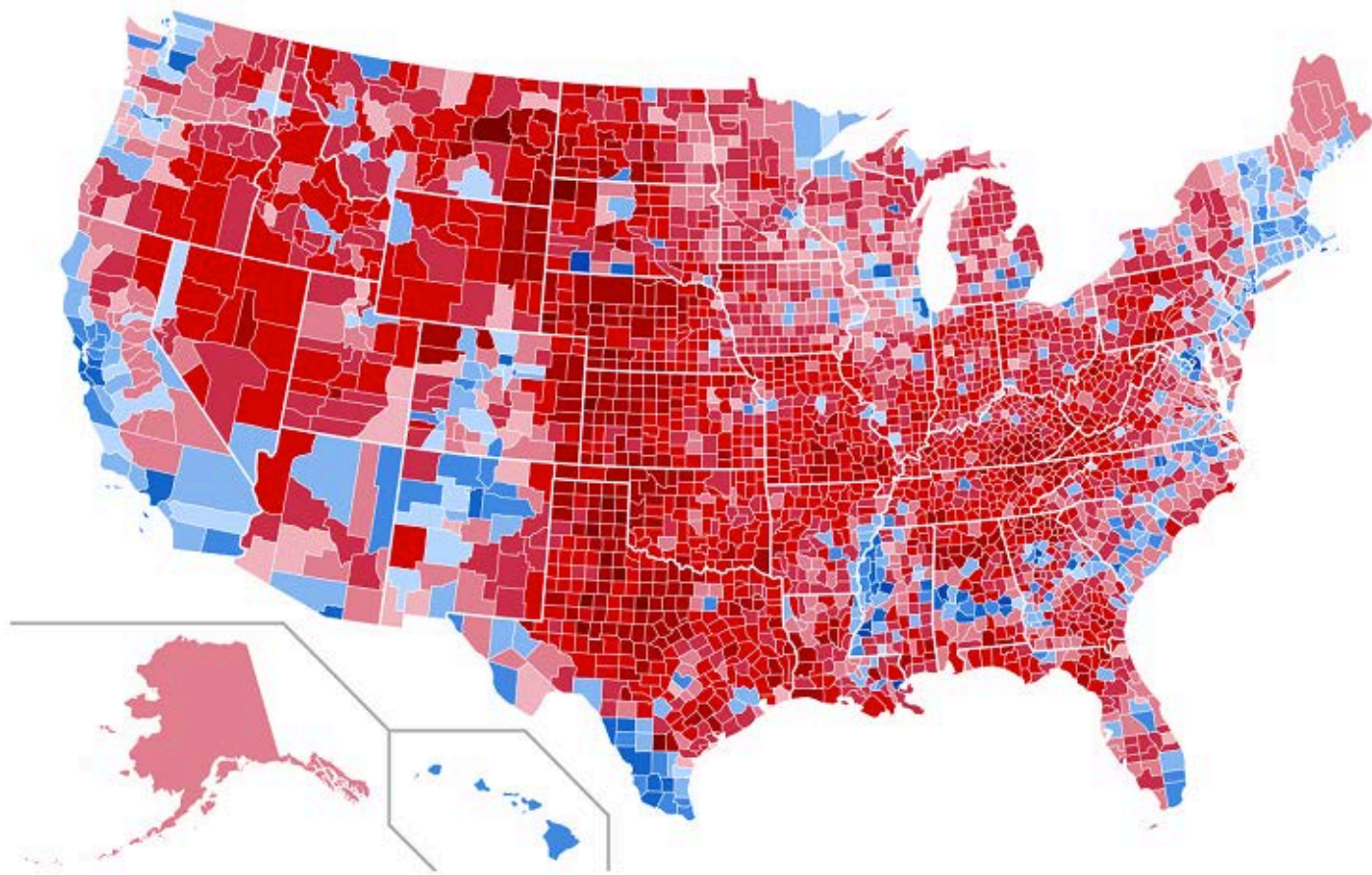
NIH Council of Councils

Friday, January 27, 2017 (1 PM)

NIH Main Campus, Building 31C, Conference Room 10

Conflicts: None

# Question for You



*Map created by Magog the Ogre via Wikimedia*

<http://brilliantmaps.com/2016-county-election-map/>



National Institutes of Health  
Office of Extramural Research

# More Questions



National Institutes of Health  
Office of Extramural Research



Unreliable research

## Trouble at the lab

*Scientists like to think of science as self-correcting. To an alarming degree, it is not*



“I SEE a train wreck looming,” warned Daniel Kahneman, an eminent psychologist, in an open letter last year. The premonition concerned

FiveThirtyEight



Politics

Sports

Science & Health

Economics

Culture

Catch up: Trump's inauguration



ILLUSTRATION BY SHOUT

## Science Isn't Broken

It's just a hell of a lot harder than we give it credit for.

By Christie Aschwanden

Filed under Scientific Method

Published Aug 19, 2015

<https://fivethirtyeight.com/features/science-isnt-broken/#part1>

<https://projects.fivethirtyeight.com/p-hacking/>



National Institutes of Health  
Office of Extramural Research

# Hack Your Way To Scientific Glory



You're a social scientist with a hunch: **The U.S. economy is affected by whether Republicans or Democrats are in office.** Try to show that a connection exists, using real data going back to 1948. For your results to be publishable in an academic journal, you'll need to prove that they are "statistically significant" by achieving a low enough p-value.

## 1 CHOOSE A POLITICAL PARTY

Republicans

Democrats

## 2 DEFINE TERMS

Which politicians do you want to include?

- ☐ Presidents
- ☒ Governors
- ☒ Senators
- ☒ Representatives

How do you want to measure economic performance?

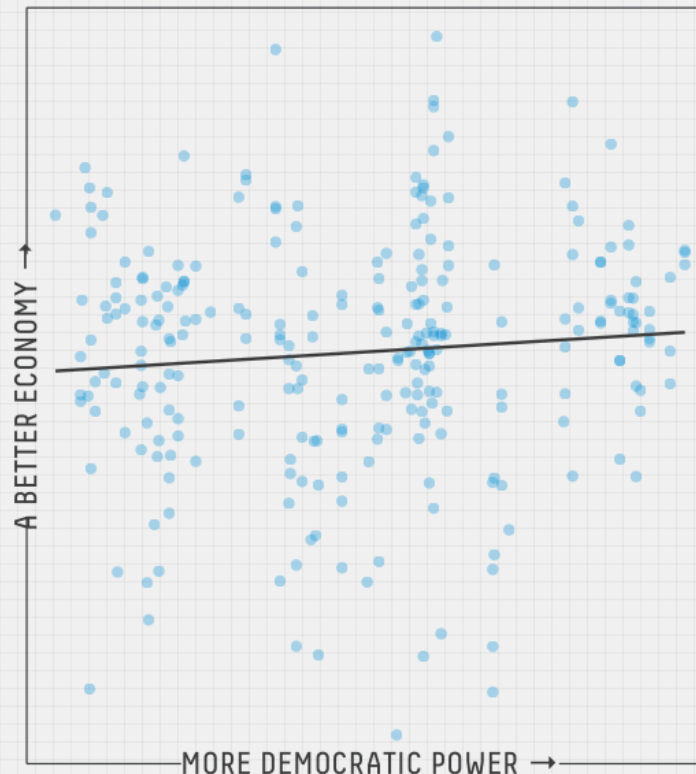
- ☐ Employment
- ☒ Inflation
- ☒ GDP
- ☒ Stock prices

Other options

- ☒ Factor in power  
Weight more powerful positions more heavily
- ☐ Exclude recessions  
Don't include economic recessions

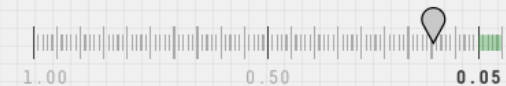
## 3 IS THERE A RELATIONSHIP?

Given how you've defined your terms, does the economy do better, worse or about the same when more Democrats are in power? Each dot below represents one month of data.



## 4 IS YOUR RESULT SIGNIFICANT?

If there were no connection between the economy and politics, what is the probability that you'd get results at least as strong as yours? That probability is your p-value, and by convention, you need a **p-value of 0.05 or less** to get published.



## Result: Unpublishable

With a p-value of **0.15**, your findings are not statistically significant. Try defining your terms differently.

If you're interested in reading real (and more rigorous) studies on the connection between politics and the economy, see the work of Larry Bartels and Alan Blinder and Mark Watson.

Data from The @unitedstates Project, National Governors Association, Bureau of Labor Statistics, Federal Reserve Bank of St. Louis and Yahoo Finance.

# Hack Your Way To Scientific Glory



You're a social scientist with a hunch: **The U.S. economy is affected by whether Republicans or Democrats are in office.** Try to show that a connection exists, using real data going back to 1948. For your results to be publishable in an academic journal, you'll need to prove that they are "statistically significant" by achieving a low enough p-value.

## 1 CHOOSE A POLITICAL PARTY

Republicans

Democrats

## 2 DEFINE TERMS

Which politicians do you want to include?

- ☒ Presidents
- ☒ Governors
- ☒ Senators
- ☒ Representatives

How do you want to measure economic performance?

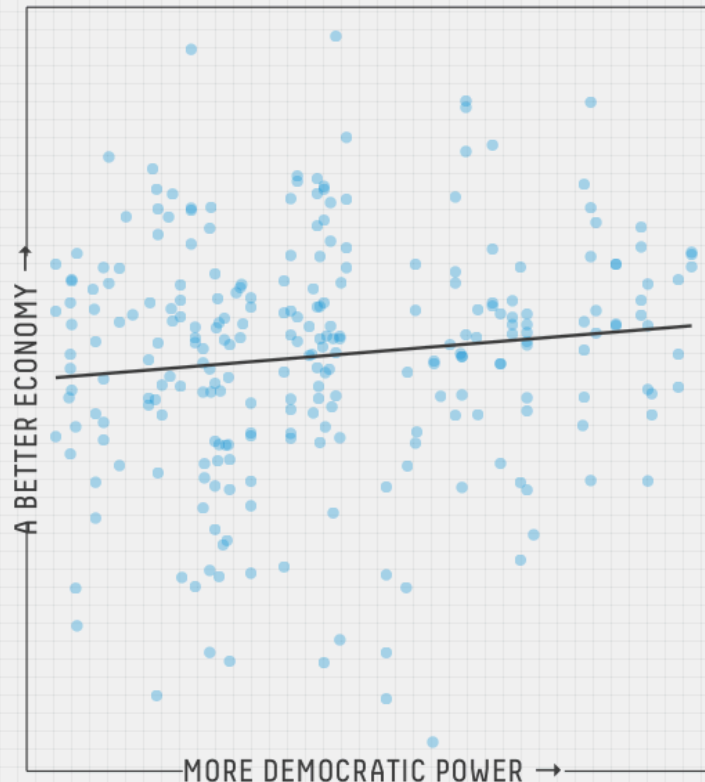
- ☐ Employment
- ☒ Inflation
- ☒ GDP
- ☒ Stock prices

Other options

- ☒ Factor in power  
Weight more powerful positions more heavily
- ☐ Exclude recessions  
Don't include economic recessions

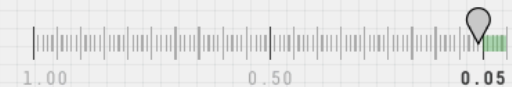
## 3 IS THERE A RELATIONSHIP?

Given how you've defined your terms, does the economy do better, worse or about the same when more Democrats are in power? Each dot below represents one month of data.



## 4 IS YOUR RESULT SIGNIFICANT?

If there were no connection between the economy and politics, what is the probability that you'd get results at least as strong as yours? That probability is your p-value, and by convention, you need a **p-value of 0.05 or less** to get published.



### Result: Almost

Your **0.06** p-value is close to the 0.05 threshold. Try tweaking your variables to see if you can push it over the line!

If you're interested in reading real (and more rigorous) studies on the connection between politics and the economy, see the work of Larry Bartels and Alan Blinder and Mark Watson.

Data from The @unitedstates Project, National Governors Association, Bureau of Labor Statistics, Federal Reserve Bank of St. Louis and Yahoo Finance.



# Hack Your Way To Scientific Glory



You're a social scientist with a hunch: **The U.S. economy is affected by whether Republicans or Democrats are in office.** Try to show that a connection exists, using real data going back to 1948. For your results to be publishable in an academic journal, you'll need to prove that they are "statistically significant" by achieving a low enough p-value.

## 1 CHOOSE A POLITICAL PARTY

Republicans

Democrats

## 2 DEFINE TERMS

Which politicians do you want to include?

- ☐ Presidents
- ☒ Governors
- ☐ Senators
- ☒ Representatives

How do you want to measure economic performance?

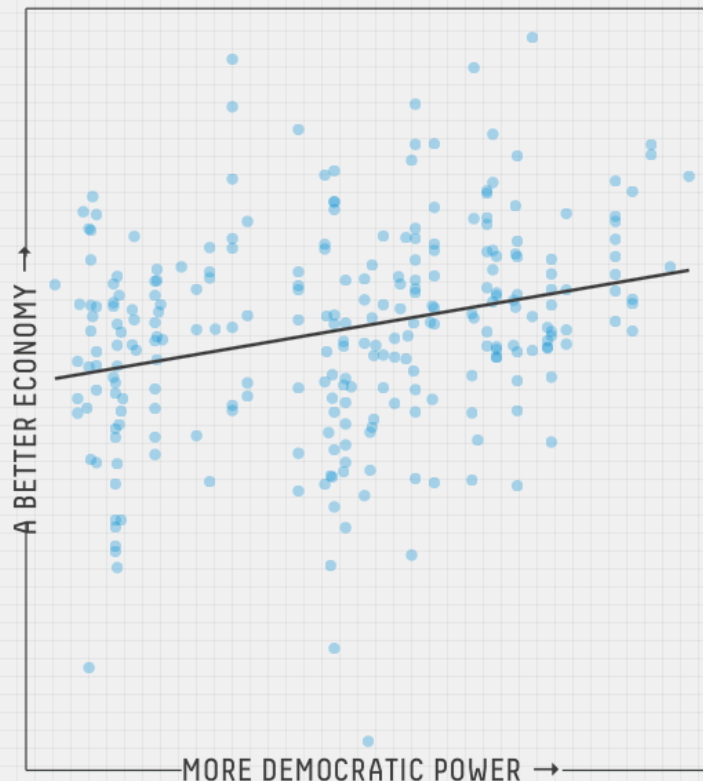
- ☐ Employment
- ☐ Inflation
- ☒ GDP
- ☒ Stock prices

Other options

- ☒ Factor in power  
Weight more powerful positions more heavily
- ☐ Exclude recessions  
Don't include economic recessions

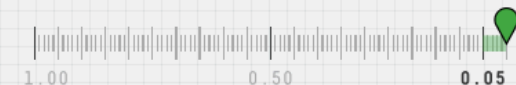
## 3 IS THERE A RELATIONSHIP?

Given how you've defined your terms, does the economy do better, worse or about the same when more Democrats are in power? Each dot below represents one month of data.



## 4 IS YOUR RESULT SIGNIFICANT?

If there were no connection between the economy and politics, what is the probability that you'd get results at least as strong as yours? That probability is your p-value, and by convention, you need a **p-value of 0.05 or less** to get published.



**Result: Publishable**

You achieved a p-value of **less than 0.01** and showed that **Democrats** have a **positive** effect on the economy. Get ready to be published!

If you're interested in reading real (and more rigorous) studies on the connection between politics and the economy, see the work of Larry Bartels and Alan Blinder and Mark Watson.

Data from The @unitedstates Project, National Governors Association, Bureau of Labor Statistics, Federal Reserve Bank of St. Louis and Yahoo Finance.



## Essay

# Why Most Published Research Findings Are False

John P.A. Ioannidis

## Lies, Damned Lies, and Medical Science

Much of what medical researchers conclude in their studies is misleading, exaggerated, or flat-out wrong. So why are doctors—to a striking extent—still drawing upon misinformation in their everyday practice? Dr. John Ioannidis has spent his career challenging his peers by exposing their bad science.

DAVID H. FREEDMAN | NOVEMBER 2010 ISSUE | TECHNOLOGY

PLoS Medicine 2005;2:e124  
Atlantic Magazine, November 2010

## Essay

# Why Most Published Research Findings Are False

John P.A. Ioannidis

“For most study designs and settings, it is more likely for a research claim to be false than true.”

- Smaller studies
- Smaller effect size
- Greater number of tested relationships
- Flexibility in designs and definitions
- Financial interests and fads

PLoS Medicine 2005;2:e124

# Likelihood of “Truth”

Study Design	Likelihood to be “True”
Well-powered RCT; no bias; 1:1 pre-study odds	85%
Meta-analysis of underpowered trials	41%
Underpowered, well-done phase 1 or 2 RCT	23%
Adequately powered observational study	20%
Underpowered observational study	12%
Discovery-oriented, massive testing	< 1%

PLoS Medicine 2005;2:e124

# PERSPECTIVE

---

doi:10.1038/nature11556

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

“... 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.**”

Landis SC et al. Nature 2012;490:187-191



Power failure: why small sample size undermines the reliability of neuroscience

*Katherine S. Button<sup>1,2</sup>, John P. A. Ioannidis<sup>3</sup>, Claire Mokrysz<sup>1</sup>, Brian A. Nosek<sup>4</sup>, Jonathan Flint<sup>5</sup>, Emma S. J. Robinson<sup>6</sup> and Marcus R. Munaf <sup>1</sup>*

	Total animals used	Required N per study		Typical N per study	
		80% power	95% power	Mean	Median
Water maze	420	134	220	22	20
Radial maze	514	68	112	24	20

“What is particularly striking is the inefficiency of a continued reliance on small sample sizes. ... Low power has an ethical dimension – unreliable research is inefficient and wasteful. This applies to both human and animal research.”

Button KS et al. Nature Reviews Neuroscience. 2013;14:365-76

# Beginning to Be Recognized



USAF/GETTY

Experiments that use only a small number of animals are common, but might not give meaningful results.

## MEDICAL RESEARCH

# UK funders demand strong statistics for animal studies

*Move addresses concerns that some experiments are not using enough animals.*

BY DANIEL CRESSEY

**R**eplace, refine, reduce: the 3 Rs of ethical animal research are widely accepted around the world. But now the message

for animal experiments. Funding applicants must now show that their work will provide statistically robust results — not just explain how it is justified and set out the ethical implications — or risk having their grant application rejected.

Sert, who works on experimental design at the National Centre for the Replacement, Refinement and Reduction (NC3Rs) of Animals in Research in London. “These animals are going to be wasted.”



**NIH** National Institutes of Health  
Office of Extramural Research

Cressey D. Nature 2015;520:271-2



## 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.

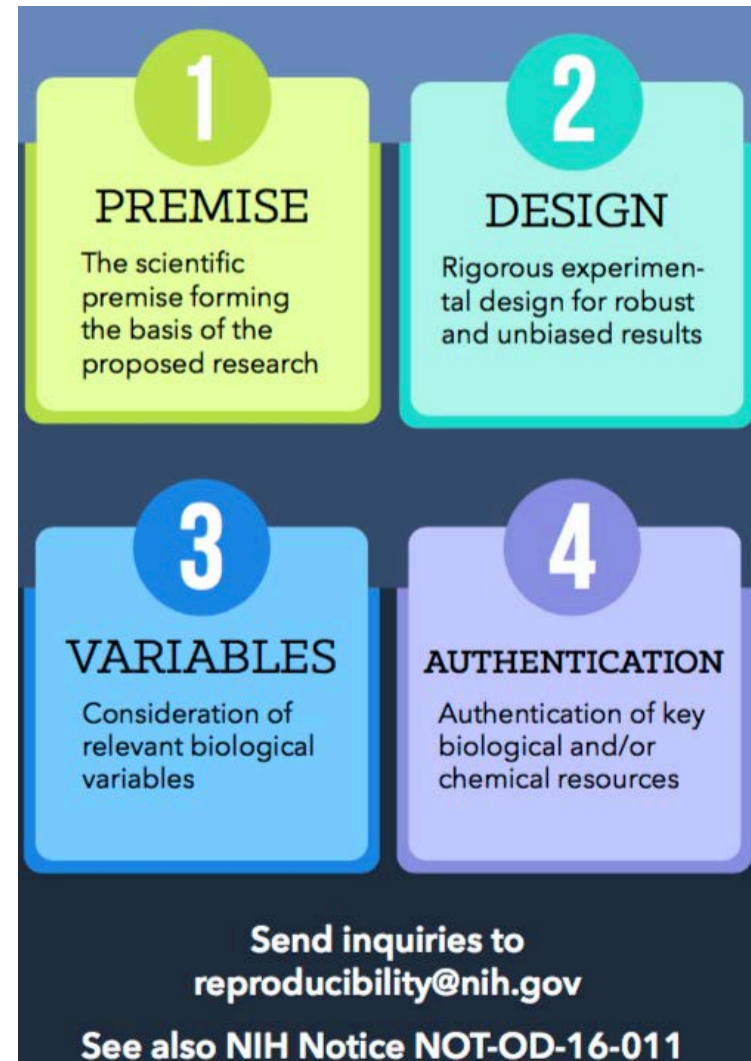
## Rigor and Reproducibility

Scientific rigor and transparency in conducting biomedical research outcomes. The information provided on this website is designed for NIH grant applications and progress reports.

### On This Page:

- [Goals](#)
- [Guidance: Rigor and Reproducibility in Grant Applications](#)
- [Resources](#)
- [News](#)
- [References](#)

<https://grants.nih.gov/reproducibility/index.htm>





# NEW GRANT GUIDELINES

what you need to know

## WHY UPDATE THE GUIDELINES?

The updates focus on four areas deemed important for enhancing rigor and transparency:

1

### PREMISE

The scientific premise forming the basis of the proposed research

2

### DESIGN

Rigorous experimental design for robust and unbiased results

3

### VARIABLES

Consideration of relevant biological variables

4

### AUTHENTICATION

Authentication of key biological and/or chemical resources

Send inquiries to  
[reproducibility@nih.gov](mailto:reproducibility@nih.gov)

See also NIH Notice NOT-OD-16-011

<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-011.html>

## WHAT ARE THE UPDATES?

### 1 UPDATES TO RESEARCH STRATEGY GUIDANCE

The research strategy is where you discuss the significance, innovation, and approach of your research plan. Let's look at an R01, for example:



Introduction to resubmission and revision applications



Specific aims



Research strategy



Commercialization plan



Biographical sketch

The new **research strategy** guidelines require that you:

- State the strengths and weakness of published research or preliminary data crucial to the support of your application
- Describe how your experimental design and methods will achieve robust and unbiased results
- Explain how biological variables, such as sex, are factored into research design and provide justification if only one sex is used

### 2 NEW ATTACHMENT FOR AUTHENTICATION OF KEY BIOLOGICAL AND/OR CHEMICAL RESOURCES

From now on, you must briefly describe methods to ensure the identity and validity of key biological and/or chemical resources used in the proposed studies.

These include, but are not limited to:



Standard laboratory reagents that are not expected to vary do not need to be included in the plan. Examples are buffers and other common biologicals or chemicals.



**DO NOT** put experimental methods or preliminary data in this section



**DO** focus on authentication and validation of key resources

### 3 NEW REVIEWER GUIDELINES

Here are the additional criteria the reviewers will be asked to use:

➔ Is there a **strong scientific premise** for the project?



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



Have the investigators presented strategies to ensure a **robust and unbiased approach**, as appropriate for the work proposed?



Reviewers will also be asked to comment on that new attachment (see Update 2)!

<https://grants.nih.gov/reproducibility/documents/grant-guideline.pdf>



National Institutes of Health  
Office of Extramural Research

# Special Steps with Clinical Trials

## VIEWPOINT

## Toward a New Era of Trust and Transparency in Clinical Trials

---

**Kathy L. Hudson, PhD**

National Institutes of Health, Bethesda, Maryland.

---

**Michael S. Lauer, MD**

National Institutes of Health, Bethesda, Maryland.

---

**Francis S. Collins, MD, PhD**

National Institutes of Health, Bethesda, Maryland.

- Dedicated FOAs
- Special review criteria
- GCP Training
- Single IRB
- Required registration, reporting
- NIH-wide oversight system

JAMA 2016 (online pub September 16, 2016)



National Institutes of Health  
Office of Extramural Research

## EDITORIAL

### *Journals unite for reproducibility*

**R**eproducibility, 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

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



*Marcia McNutt*  
*Editor-in-Chief*  
*Science Journals*

## Enhancing Research Reproducibility:

Recommendations from the  
Federation of American Societies for Experimental Biology



# FASEB

Federation of American Societies  
for Experimental Biology

Science. 2014;346:679



National Institutes of Health  
Office of Extramural Research





Courtesy of Linda A. Cicero, Stanford University

whose results are irreproducible.

## LECTURE IV

### Statistical Proof and the Problem of Irreproducibility

*Friday, January 6, 2017, Starting at 4:00 p.m. Imperial Ballroom A, Marquis Level, Marriott Marquis Atlanta*

**Susan Holmes**, Stanford University

Data currently generated in the fields of ecology, medicine, climate science and neuroscience often contain tens of thousands of measured variables. Statistical analyses can result in publications

[http://jointmathematicsm meetings.org/meetings/national/jmm2017/2180\\_invspeakers#holm2](http://jointmathematicsm meetings.org/meetings/national/jmm2017/2180_invspeakers#holm2)  
 Thanks to Dr. Jonathan Rosenberg



- Publish protocol, code, and data
- Divide exploration, confirmation
- Honesty
  - Secret data dredging
  - Data dropping

[http://jointmathematicsm meetings.org/meetings/national/jmm2017/2180\\_invspeakers#holm2](http://jointmathematicsm meetings.org/meetings/national/jmm2017/2180_invspeakers#holm2)



# Appreciate Your Help With Communications

## PERSPECTIVES



GBSI™

Global Biological Standards Institute™

ABOUT GBSI

OUR W

## Antibody Validation: Standards, Policies, and Practices

September 25, 2016 - September 27, 2016

Asilomar Conference Grounds

*The FASEB Journal* • Life Sciences Forum

## Studying both sexes: a guiding principle for biomedicine

**Janine Austin Clayton**

Office of Research on Women's Health, National Institutes of Health, Bethesda, Maryland, USA

## CELL BIOLOGY

## Fixing problems with cell lines

Technologies and policies can improve authentication

By Jon R. Lorsch<sup>1\*</sup>, Francis S. Collins<sup>2</sup>,  
Jennifer Lippincott-Schwartz<sup>3,4</sup>

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

For example, studies using just two misidentified cell lines were included in three grants funded by the U.S. National Institutes of

Science 2014;346:1452-3

<https://www.gbsi.org/event/asilomar/>

FASEBJ. 2016;30:519-24



National Institutes of Health  
Office of Extramural Research

- Deep-seated cognitive biases
  - Belief in small numbers, neglect of priors
  - Missing regression to the mean, confounders
  - Principles of experimental design, reporting
  - The problem of multiplicity
- Multi-stakeholder strategy key
- Evaluation ongoing

Kahneman D. Farrar, Straus, and Giroux. 2011